

**Tilburg University**

## **Estimation of Spatial Sample Selection Models**

Rabovic, Renata; Cizek, Pavel

*Publication date:*  
2016

*Document Version*  
Early version, also known as pre-print

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Rabovic, R., & Cizek, P. (2016). *Estimation of Spatial Sample Selection Models: A Partial Maximum Likelihood Approach*. (CentER Discussion Paper; Vol. 2016-013). CentER, Center for Economic Research.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

No. 2016-013

**ESTIMATION OF SPATIAL SAMPLE SELECTION MODELS:  
A PARTIAL MAXIMUM LIKELIHOOD APPROACH**

By

Renata Rabovič, Pavel Čížek

31 March 2016

ISSN 0924-7815  
ISSN 2213-9532

# Estimation of Spatial Sample Selection Models: A Partial Maximum Likelihood Approach

Renata Rabovič\*      Pavel Čížek†

March 23, 2016

## Abstract

To analyze data obtained by non-random sampling in the presence of cross-sectional dependence, estimation of a sample selection model with a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations is considered. Since there is no estimation framework for the spatial lag model and the existing estimators for the spatial error model are either computationally demanding or have poor small sample properties, we suggest to estimate these models by the partial maximum likelihood estimator, following Wang et al. (2013)'s framework for a spatial error probit model. We show that the estimator is consistent and asymptotically normally distributed. To facilitate easy and precise estimation of the variance matrix without requiring the spatial stationarity of errors, we propose the parametric bootstrap method. Monte Carlo simulations demonstrate the advantages of the estimators.

*JEL codes:* C13, C31, C34

*Keywords:* asymptotic distribution, maximum likelihood, near epoch dependence, sample selection model, spatial autoregressive model

---

\*Corresponding author. CentER, Dept. of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands. **Email:** R.Rabovic@tilburguniversity.edu, Phone: +31 13 466 4412.

†CentER, Dept. of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands. **Email:** P.Cizek@tilburguniversity.edu, Phone: +31 13 466 8723.

# 1 Introduction

The assumption about independent observations is often not met even in the analysis of cross-sectional data. Since cross-sectional dependence can be captured by a certain spatial or economic ordering in many economic applications, spatial models have become an extensively used tool in applied econometrics. In this paper, we propose spatial extensions of sample selection models. We introduce spatial dependence into a sample selection model via a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations. To our best knowledge, this is the first paper which analyzes a sample selection model with a spatial lag of a latent dependent variable, facilitating easy estimation in applications such as peer effects in education with non-randomly missing data (see Section 2 for more details). A spatial sample selection model with a spatial error, which can be used, for instance, in agricultural yield studies, has been analyzed before, but the proposed estimators are either computationally demanding or they do not have desirable small sample performance.

The computational difficulties in the spatial sample selection models stem from the (spatially) correlated errors: their joint density function cannot be expressed as a product of the density functions for each observation, and the full maximum likelihood estimator (MLE) becomes computationally very demanding as it involves high dimensional integration. It is possible to overcome this obstacle by using the heteroskedastic maximum likelihood estimator (HMLE), which takes into account only heteroskedasticity stemming from spatial correlation while neglecting the corresponding spatial autocorrelation to obtain consistent but inefficient estimates.<sup>1</sup> Based on this idea, Flores-Lagunes and Schnier (2012) in the context of a sample selection model with a spatial error in both the selection and outcome equations proposed to use the generalized method of moments (GMM) estimator.<sup>2</sup> The estimator however has poor small sample properties (see Section 4 in their paper and Section 5). Doğan and Taşpınar (2014) suggest to estimate the same model using the Markov chain Monte Carlo approach in the context of Bayesian estimation, whereas earlier McMillen (1995) suggested to use the Expectation Maximization algorithm. Both of these methods are however computationally demanding in larger samples, and moreover, a rigorous theory is not developed for either of them.

In the closely related context of binary choice models with spatial errors, Wang et al. (2013) therefore suggested an intermediate approach between the full MLE and HMLE that is based on the idea that all observations are divided into clusters of two observations and the dependence within clusters is taken into

---

<sup>1</sup>Poirier and Ruud (1988) developed the result under fairly general conditions for a probit model with serial correlation in a time series setting, whereas Robinson (1982) established the same result for a Tobit model.

<sup>2</sup>For empirical studies that use the estimator suggested by Flores-Lagunes and Schnier (2012), see Section 5 of Flores-Lagunes and Schnier (2012), Mukherjee and Singer (2010), and Ward et al. (2014).

account, whereas the dependence between clusters is not employed in the estimation. This approach avoids the computationally demanding full MLE, while it facilitates the estimation of the spatial error structure by taking the correlation within clusters into account. Wang et al. (2013) apply the partial maximum likelihood estimator (PMLE) to a spatial error probit model. In this paper, the PMLE approach is generalized to sample selection models with a spatial lag of a latent dependent variable or a spatial error and their special cases.

Since the special cases of the considered sample selection models include probit and Tobit models (see Section 2), this paper also extends Wang et al. (2013)'s results to the probit and Tobit models with a spatial lag of the latent dependent variable and to the Tobit model with a spatial error.<sup>3</sup> We analyze the asymptotic properties of the proposed PMLE using the near epoch dependent random fields framework introduced by Jenish and Prucha (2012). Note that the asymptotic results derived for a spatial error probit model in Wang et al. (2013) cannot be directly applied to our models because the structure of the spatial sample selection models is more complicated and requires additional treatment. For example, the uniform  $L_p$ -boundedness of the (bivariate) likelihood scores cannot be established by simply assuming that the support of exogeneous regressors is bounded since the observed dependent variables also enter the cumulative distribution function of the bivariate normal distribution. Moreover, Wang et al. (2013) base their analysis on  $\alpha$ -mixing processes and make assumptions about dependence based on the observed responses instead of deriving more primitive conditions.<sup>4</sup> They also impose a strong assumption on the expansion speed of the sampling region<sup>5</sup> and suggest to estimate the variance matrix of the proposed estimator based on the approach proposed by Conley (1999), who explicitly models the sampling process from a regular lattice and assumes that the data generating process is strongly spatially stationary. This assumption is in general not satisfied, for example, for the Cliff-Ord type models (see Kelejian and Prucha, 2007, for a further discussion). We relax these assumptions and suggest to estimate the asymptotic variance matrix using parametric bootstrap.

The paper is organized as follows. In Section 2, the sample selection models are defined, whereas the PMLE is introduced in Section 3. In Section 4, its consistency and asymptotic normality are established, and the estimator of the asymptotic variance matrix is discussed. In Section 5, we study the finite sample properties, while Section 6 concludes. Proofs are provided in Appendices.

It proves helpful to introduce the following notation. Let  $A_n$ ,  $n \in \mathbb{N}$ , be some matrix indexed by  $n$ ; we denote the  $ij$ th element, the  $i$ th row, and the  $j$ th column of  $A_n$  by  $A_{ij,n}$ ,  $A_{i,n}$ , and  $A_{\cdot j,n}$ , respectively.

---

<sup>3</sup>The Tobit model with a spatial lag of the observed dependent variable has been recently analyzed by Xu and Lee (2015a).

<sup>4</sup>See conditions (vii) and (i) of Theorems 1 and 2 by Wang et al. (2013), respectively.

<sup>5</sup>See condition (ii) of Theorem 2 by Wang et al. (2013).

Similarly, if  $v_n$  is a vector, then  $v_{i,n}$  denotes the  $i$ th element of  $v_n$ . (The same notation applies for vectors and matrices that are not indexed by  $n$ .) Further, let  $g = (i, j)'$  and  $\dot{g} = (k, l)'$ . Then  $A_{g,n} = (A'_{i,n}, A'_{j,n})'$  with its  $qm$ th element,  $q$ th row, and  $m$ th column denoted by  $A_{gqm,n}$ ,  $A_{gq,n}$ , and  $A_{g,m,n}$ , respectively;  $A_{g\dot{g},n} = [A_{ik,n} \ A_{il,n}; \ A_{jk,n} \ A_{jl,n}]$  and  $A_{g,n} = A_{gg,n}$  with  $A_{g,n} = [A_{g11,n} \ A_{g12,n}; \ A_{g21,n} \ A_{g22,n}]$ . Similarly,  $v_{g,n} = (v_{i,n}, v_{j,n})'$  with its  $q$ th element denoted by  $v_{gq,n}$ . Furthermore, for any random vector  $Y$ , let  $\|Y\|_p = [E\|Y\|^p]^{1/p}$ ,  $p \geq 1$ , denote its  $L_p$ -norm, where  $\|\cdot\|$  is the Euclidean norm. For an  $n \times n$  matrix  $A$ , the Euclidean, row sum, and column sum matrix norms are defined as  $\|A\| = \left(\sum_{i=1}^n \sum_{j=1}^n |A_{ij}|^2\right)^{1/2}$ ,  $\|A\|_\infty = \max_{i=1,\dots,n} \sum_{j=1}^n |A_{ij}|$ , and  $\|A\|_1 = \max_{j=1,\dots,n} \sum_{i=1}^n |A_{ij}|$ , respectively.<sup>6</sup> Note that these norms are sub-multiplicative:  $\|AB\|_a \leq \|A\|_a \|B\|_a$ , where  $\|\cdot\|_a$  denotes one of the mentioned norms.

## 2 Model

To define the sample selection model, consider first the following latent selection ( $s$ ) and outcome ( $o$ ) equations with spatial lags of the latent dependent variable:

$$\begin{aligned} y_{i,n}^{*s} &= \lambda_0^s W_{i,n}^s y_n^{*s} + X_{i,n}^s \beta_0^s + u_{i,n}^s \\ y_{i,n}^{*o} &= \lambda_0^o W_{i,n}^o y_n^{*o} + X_{i,n}^o \beta_0^o + u_{i,n}^o \end{aligned} \tag{1}$$

for  $i = 1, \dots, 2n$ ,<sup>7</sup> where  $2n$  represents the actual sample size and  $n$  serves as the sample-size index,  $y_{i,n}^{*s}$  and  $y_{i,n}^{*o}$  are latent variables,  $X_{i,n}^s$  and  $X_{i,n}^o$  are  $1 \times L^s$  and  $1 \times L^o$  dimensional vectors of exogenous variables, and  $u_{i,n}^s$  and  $u_{i,n}^o$  are the error terms for the selection and outcome equations, respectively; the corresponding vectors and matrices of all observations are denoted by  $y_n^{*s} = (y_{i,n}^{*s})_{i=1}^{2n}$ ,  $W_n^s = (W_{i,n}^s)_{i=1}^{2n}$ ,  $X_n^s = (X_{i,n}^s)_{i=1}^{2n}$ ,  $u_n^s = (u_{i,n}^s)_{i=1}^{2n}$  and analogously for the outcome equation. The spatial nonstochastic weight matrices  $W_n^s$  and  $W_n^o$  are assumed to be known, contain nonnegative elements, and have zero elements on the main diagonal. For example, the elements of  $W_n^s$  and  $W_n^o$  can be indirectly proportional to the strength of an economic relationship or distance between two observations, or they can be equal to 0 or 1, indicating unrelated or related (neighboring) observations (e.g., see LeSage and Pace, 2009). If the  $ij$ th element of the spatial weight matrix is nonzero, there is a direct dependence between the latent variables of observations  $i$  and  $j$ . If the  $ij$ th element of the spatial weight matrix is zero, it does not mean that observations  $i$  and  $j$  are independent because there might exist an observation  $k$  that has an effect on the latent variables of both observations  $i$  and  $j$ .

<sup>6</sup>See Horn and Johnson (1985, pp. 291, 294-295) for more details.

<sup>7</sup>For notational convenience, we assume that the number of observations is even.

The relation between the observed outcomes and latent variables in (1) is defined as  $y_{i,n}^s = \mathbb{1}(y_{i,n}^{*s} > 0)$  and  $y_{i,n}^o = y_{i,n}^s y_{i,n}^{*o}$ , so that the selection equation determines which cases are observed, while the outcome equation determines the magnitude of the observed responses. (In general,  $y_{i,n}^o$  is missing for observation  $i$  rather than being zero if  $y_{i,n}^s = 0$ , but the definition is made for the simplicity of notation similarly to Chen and Zhou, 2010, among others and does not affect the likelihood function.) This version of a sample selection model is chosen because it is used in many empirical applications and includes other important models. For example, under normality of errors, modelling just  $y_{i,n}^s$  leads to probit, and taking equations in (1) identical results in Tobit. Generalizations to other sample selection models might also be considered. For instance, a binary sample selection model with  $y_{i,n}^s = \mathbb{1}(y_{i,n}^{*s} > 0)$  and  $y_{i,n}^o = y_{i,n}^s \mathbb{1}(y_{i,n}^{*o} > 0)$  or a model with a Tobit selection equation defined as  $y_{i,n}^s = \max\{0, y_{i,n}^{*s}\}$  and  $y_{i,n}^o = \mathbb{1}(y_{i,n}^s > 0) y_{i,n}^{*o}$ . Finally, the latent model (1) can be easily adapted to include spatial errors instead of spatial lags:

$$\begin{aligned} y_{i,n}^{*s} &= X_{i,n}^s \beta_0^s + \varepsilon_{i,n}^s(\lambda_0^s) \\ y_{i,n}^{*o} &= X_{i,n}^o \beta_0^o + \varepsilon_{i,n}^o(\lambda_0^o), \end{aligned} \tag{2}$$

where  $\varepsilon_{i,n}^s(\lambda_0^s) = \lambda_0^s W_{i,n}^s \varepsilon_n^s(\lambda_0^s) + u_{i,n}^s$  and  $\varepsilon_{i,n}^o(\lambda_0^o) = \lambda_0^o W_{i,n}^o \varepsilon_n^o(\lambda_0^o) + u_{i,n}^o$  similarly to  $y_{i,n}^{*s}$  and  $y_{i,n}^{*o}$  in (1) with  $\varepsilon_n^s(\lambda_0^s) = (\varepsilon_{i,n}^s(\lambda_0^s))_{i=1}^{2n}$  and  $\varepsilon_n^o(\lambda_0^o) = (\varepsilon_{i,n}^o(\lambda_0^o))_{i=1}^{2n}$ ; the observed variables are defined in the same way as before. The results presented in the paper are also derived and hold for this sample selection model with spatial errors. Adjusting for different spatial error structures such as  $\varepsilon_{i,n}^s(\lambda_0^s) = \lambda_0^s W_{i,n}^s u_n^s + u_{i,n}^s$  with  $u_n^s = (u_{i,n}^s)_{i=1}^{2n}$  is straightforward.

An important feature of the latent model in (1) is that spatial lags of the latent instead of observed variables are included in both the selection and outcome equations. For the outcome equation, it is true that  $y_{j,n}^o = y_{j,n}^{*o}$  if  $y_{j,n}^s = 1$ . Thus, the outcome equations with a lag of the latent variable and a lag of the observed variable differ primarily by the presence of  $y_{j,n}^{*o}$  with  $y_{j,n}^s = 0$  on the right hand side of the equation. Note also that  $\varepsilon_{i,n}^o(\lambda_0^o)$  in (2) in general depends on  $u_{j,n}^o$  with  $y_{j,n}^s = 0$ . For the selection equation,  $y_{j,n}^s$  and  $y_{j,n}^{*s}$  differ though, and plausibility of the model depends on whether only individuals' decisions or also their motives are observable to others. By means of two empirical examples we will illustrate, where models (1) and (2) can be a plausible specification; see also a related discussion in Qu and Lee (2012) for the censored model.

**Example 1** (Peer effects in education with non-randomly missing data). In education studies of peer effects without missing data, the outcome equation in (1) is estimated, where the dependent variable is typically a test score (e.g, Lin, 2010; Duflo et al., 2011). In some cases, test scores are unfortunately not known for all

the students who took a course. For example, Booij et al. (2015) found that, in the department of economics and business of the University of Amsterdam, only 46% of students write all the first-year exams during the first year of study. A sample selection problem arises if a student's decision to take an exam and his score depend on the student's abilities to succeed in the subject. Such a situation can be handled using model in (1) with  $y_{i,n}^s$  and  $y_{i,n}^{*o}$  being a student's decision to take an exam and his (potential) score from that exam, respectively. These decisions exhibit additionally cross-sectional dependence as the student's decision to take an exam might depend on his peers' attitudes towards taking the exam; the decision cannot depend though on his peers' decisions because they are made simultaneously by actually coming to the exam. Thus, a lag of the latent rather than observed dependent variable should be included in the selection equation. Moreover, since students who do not take the exam still affect their peers and are affected by their peers, for example, by solving assignments together and by attending the same tutorial classes, a lag of the latent dependent variable should be included in the outcome equation instead of the observed variable "test score" known only for the participants of the exam.

Model (1) can also be used to study cases when a missing data problem arises due to non-responsiveness to a survey. Consider, for example, the National Longitudinal Study of Adolescent Health data, which has been extensively used to study peer effects in education (e.g., Calvò-Armengol et al., 2009; Lin, 2010). In this data, Hoshino (2016) found that information on GPA is missing for 11% of the respondents (after taking into account missing data on the exogenous observations used in his study). It might be the case that unobserved abilities of a student affect both his decision to reveal his GPA and his GPA itself. Model (1) can thus be used to account for this kind of sample selection, where a lag of a latent variable is included in the outcome equation and is not included in the selection equation as the students filled-in the questionnaire independently.

**Example 2** (Agricultural yield). Ward et al. (2014) apply the GMM estimator proposed by Flores-Lagunes and Schnier (2012) to estimate a cereal yield response function taking into account potential sample selection bias due to farmers' endogenous decision about whether to plant cereals. Flores-Lagunes and Schnier (2012) consider the latent model in (2) with spatially correlated errors and  $W_n^s = W_n^o$  (see, for instance, equations (1) and (2) in their paper), but for the estimation, observations with  $y_{i,n}^s = 0$  are omitted from the weight matrix in the outcome equation (see footnote 16 of their paper). The estimator is thus inconsistent with the model. Ward et al. (2014) overcame this issue by choosing  $W_n^o \neq W_n^s$  in such a way that  $W_{ij,n}^o$  can have positive values only if  $y_{i,n}^s = y_{j,n}^s = 1$ . In this case, the weight matrix depends on potentially endogenous farmers' decisions whether to plant cereals. This approach however requires further research as neither PMLE nor the



GMM estimator proposed by Flores-Lagunes and Schnier (2012) are designed for the cases when the weight matrix in the outcome equation depends on the outcomes in the selection equation. On the one hand, if the correlation among unobservables in the outcome equation is driven by production technology or knowledge spillovers, then the farmers who decided not to plant a field do not likely have a lot of influence on those who decided to plant a field, and the weight matrix  $W_n^o \neq W_n^s$  considered in Ward et al. (2014) should be chosen. On the other hand, if the correlation among unobservables is mainly driven by unobserved geographical and meteorological characteristics, then both planted and not planted fields are affected similarly if they are close to each other. Since the unobserved geographical and meteorological characteristics affect both the decision to plant a field and a cereal yield response function, a nonstochastic weight matrix which captures the closeness of fields can be used, and the specification in (2) with  $W_n^o = W_n^s$  should be considered.

Model (1) can be written in a reduced form, provided that the respective inverses exist, as

$$\begin{aligned} y_n^{*s} &= S_n^s(\lambda_0^s) X_n^s \beta_0^s + \varepsilon_n^s(\lambda_0^s) \\ y_n^{*o} &= S_n^o(\lambda_0^o) X_n^o \beta_0^o + \varepsilon_n^o(\lambda_0^o), \end{aligned} \tag{3}$$

where the observed responses  $y_{i,n}^s = \mathbb{1}(y_{i,n}^{*s} > 0)$  and  $y_{i,n}^o = y_{i,n}^s y_{i,n}^{*o}$ , and for  $b \in \{s, o\}$ , matrices  $S_n^b(\lambda) = (I_{2n} - \lambda W_n^b)^{-1}$  and errors  $\varepsilon_n^b(\lambda) = S_n^b(\lambda) u_n^b$ . These definitions of  $\varepsilon_n^s(\lambda_0^s)$  and  $\varepsilon_n^o(\lambda_0^o)$  are equivalent to those in the spatial error model (2), and models (2) and (3) thus differ only by the presence of  $S_n^s(\lambda_0^s)$  and  $S_n^o(\lambda_0^o)$  in the latter model.<sup>8</sup> The spatial weight matrices in the original and reduced form models have to satisfy the following assumption.

**Assumption 1.** (i) The matrices  $I_{2n} - \lambda^s W_n^s$  and  $I_{2n} - \lambda^o W_n^o$  are nonsingular for all  $(\lambda^s, \lambda^o)' \in \Lambda$ , where  $\Lambda$  is the space of the spatial parameters. (ii) The row and column sum matrix norms of matrices  $W_n^s$ ,  $W_n^o$ ,  $S_n^s(\lambda^s)$ , and  $S_n^o(\lambda^o)$  are bounded uniformly in  $n \in \mathbb{N}$  and  $(\lambda^s, \lambda^o)' \in \Lambda$ .

The first condition implies that there is a unique solution to  $y_n^{*s}$  and  $y_n^{*o}$  in (1) as well as to  $\varepsilon_n^s(\lambda_0^s)$  and  $\varepsilon_n^o(\lambda_0^o)$  in (2). Since there is no natural parameter space for spatial parameters, this condition is usually ensured by normalizing spatial weight matrices and bounding the parameter space. In applications, the weight matrices are typically normalized in such a way that the sum of each row is equal to 1 and the parameter space of  $(\lambda^s, \lambda^o)$  is chosen to be  $(-1, 1) \times (-1, 1)$ . However, if there is no theoretical reason for the row normalization, this might lead to misspecification. Kelejian and Prucha (2010) instead suggest to normalize the weight matrices by their largest absolute eigenvalues. The second condition restricts dependence to a

---

<sup>8</sup>For the simplicity of notation, we do not consider models with both spatial lags and spatial errors in both the selection and outcome equations. These models can however be analyzed in a similar way as the spatial lag model.

manageable degree. This is a classical assumption in the spatial econometrics literature (e.g., see Kelejian and Prucha, 1998, 1999, 2010).

### 3 Partial Maximum Likelihood Estimator

The (partial) maximum likelihood estimator requires a parametric specification of the distribution of the error terms. Although we could consider a general elliptically contoured distribution of  $(u_{i,n}^s, u_{i,n}^o)'$ , we restrict our attention to the Gaussian case as it turns out to be not only the most frequently used one, but also the most complicated one (relative to heavier-tailed distributions) due to the necessity to study and bound the moments of the logarithm of the bivariate normal cumulate distribution function and their derivatives. Let  $\theta = (\beta^{s'}, \beta^{o'}, \lambda^s, \lambda^o, \rho, \sigma^2)'$ .

**Assumption 2.** (i) The error terms  $(u_{i,n}^s, u_{i,n}^o)' \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma(\theta_0))$ , where  $\Sigma(\theta_0) = \begin{bmatrix} 1 & \rho_0\sigma_0 \\ \rho_0\sigma_0 & \sigma_0^2 \end{bmatrix}$  is a positive definite matrix. (ii)  $(X_n^s, X_n^o)$  and  $(u_n^s, u_n^o)$  are independent. (iii)  $X_{i,n}^s, X_{i,n}^o, W_{i,n}^s$ , and  $W_{i,n}^o$  are always observed.

Assumption 2(i) is strong but standard in the literature that analyzes parametric sample selection models (see Heckman, 1974, 1979). The variance of  $u_{i,n}^s$  is normalized to 1 in order to ensure identification. The correlation coefficient  $\rho_0$  controls the selection bias; if  $\rho_0$  is zero, the outcome equation can be estimated independently of the selection equation. Even in that case, standard estimators for spatial linear models, for example, MLE (see Lee, 2004) or GMM (see Kelejian and Prucha, 1998), cannot be applied in order to estimate the outcome equation in (1) for a subsample of observations with  $y_{i,n}^s = 1$  because  $y_{j,n}^{*o}$  on the right hand side of the outcome equation is missing if  $y_{j,n}^s = 0$ . If data are missing at random, the methods developed by Wang and Lee (2013) for estimation of spatial autoregressive models are applicable to model (1) and a version of the MLE estimator<sup>9</sup> can be applied to the outcome equation in (2) for a subsample of observations with  $y_{i,n}^s = 1$ . Neither method is applicable if  $\rho_0 \neq 0$  though.

Further, in the standard sample selection model, it is assumed that  $(X_{i,n}^s, X_{i,n}^o)$  and  $(u_{i,n}^s, u_{i,n}^o)$  are independent. Due to spatially correlated errors  $\varepsilon_{i,n}^s(\lambda^s)$  and  $\varepsilon_{i,n}^o(\lambda^o)$ , which are present in both the spatial lag and spatial error models in (3) and (2), respectively, we need to make an assumption that the exogenous variables and the error terms are mutually independent as in Assumption 2(ii). Finally, Assumption 2(iii) states that the exogenous variables and spatial weights have to be observed even for missing observations. The observability of the exogenous variables in the outcome equation for missing observations is not required

<sup>9</sup>If  $u_{i,n}^o | X_{i,n}^o \sim \mathcal{N}(0, \sigma_0^2)$ , then  $y_{i,n}^{*o} | X_{i,n}^o \sim \mathcal{N}(X_{i,n}^o \beta_0^o, S_{i,n}^o(\lambda_0^o) S_{i,n}^o(\lambda_0^o) \sigma_0^2)$ . If the dependent variable is missing randomly,  $y_{i,n}^o | X_{i,n}^o$  follows the same distribution.

neither in the standard parametric sample selection model nor in model (2) with spatial errors. If the model contains a spatial lag, the partial maximum likelihood estimator is however based on the reduced form in (3), which requires full observability. This assumption is not too strong since  $X_{i,n}^s$  contains variables in  $X_{i,n}^o$  in many empirical applications (e.g., Buchinsky, 1998; Vella, 1998; Sharma et al., 2013). The assumption about observability of spatial weights is not very restrictive either, at least if the spatial weights are based on distances between observations, which are typically not difficult to obtain even for missing observations.

Due to the spatial dependence, the error terms  $\varepsilon_{i,n}^s(\lambda^s)$  and  $\varepsilon_{i,n}^o(\lambda^o)$  in latent models (3) and (2) are heteroskedastic and cross-correlated. Hence, the full MLE is computationally demanding in this setting. Based on the idea introduced by Wang et al. (2013), we therefore suggest to estimate the models by applying the partial maximum likelihood estimator. Specifically, we divide  $2n$  observations into  $n$  mutually disjoint pairs based on the idea that the internal correlation between two observations in a pair is more important than the external correlation with observations in the other pairs, at least if observations within a pair are “close” to each other. If only very weakly correlated observations are paired, the estimator will be similar to HMLE and there will be no gains from forming the pairs. The way how the observations are paired thus has an effect on the estimation precision, and it is desirable to group observations in such a way that the variance of the estimator is minimized. Given that the asymptotic variance is a function of unknown parameters, a two-step procedure might be considered, where the initial estimates based on some primitive grouping are used to construct an optimal grouping. It is unfortunately very difficult, if not impossible, to construct the asymptotic distribution of such a two-step estimator because the grouping becomes data dependent. Moreover, it is not clear how to obtain an optimal grouping practically as it would involve huge computational costs unless very rough approximations of the asymptotic variance are used. For these reasons, we suggest to group observations based on deterministic variables that potentially capture the strength of dependence between observations, for example, the Euclidean distance between observations (see Section 5.1 for details). As discussed in Wang et al. (2013), it is also possible to try a finite number of different grouping schemes and to choose the one which delivers the smallest standard errors.

Let a grouping of observations be described by an index set  $\mathcal{G}_n$  containing  $n$  pairs  $g = (i, j)'$  of observations  $i$  and  $j$ ;  $\cup_{g \in \mathcal{G}_n} \{g_1, g_2\} = \{1, \dots, 2n\}$ . Let  $y_{g,n}^{*s} = (y_{g1,n}^{*s}, y_{g2,n}^{*s})'$  and  $y_{g,n}^{*o} = (y_{g1,n}^{*o}, y_{g2,n}^{*o})'$  be 2-dimensional vectors of latent variables in a group  $g \in \mathcal{G}_n$ . The latent processes for a group  $g$  from the reduced form in

(3) can be then written as

$$\begin{aligned} y_{g,n}^{*s} &= S_{g,n}^s(\lambda_0^s) X_n^s \beta_0^s + \varepsilon_{g,n}^s(\lambda_0^s) \\ y_{g,n}^{*o} &= S_{g,n}^o(\lambda_0^o) X_n^o \beta_0^o + \varepsilon_{g,n}^o(\lambda_0^o) \end{aligned}$$

with observed responses  $y_{g,n}^s$  and  $y_{g,n}^o$ , where all variables are defined in the same way as in Section 2 except that now they are defined for a group  $g$  instead of an individual  $i$ ; that is,  $S_{g,n}^s(\lambda_0^s) = (S_{g1,n}^{s'}(\lambda_0^s), S_{g2,n}^{s'}(\lambda_0^s))'$ ,  $\varepsilon_{g,n}^s(\lambda_0^s) = (\varepsilon_{g1,n}^s(\lambda_0^s), \varepsilon_{g2,n}^s(\lambda_0^s))'$ ,  $y_{g,n}^s = (y_{g1,n}^s, y_{g2,n}^s)'$ , and so on. The grouped spatial error model can be defined analogously.

Before constructing the log-likelihood function  $Q_n(\theta)$  and its population counterpart  $Q_0(\theta)$ , note that the log-likelihood function will be composed of four parts because there are four scenarios: one observation in a pair is missing ( $y_{g1,n}^s = 1$  and  $y_{g2,n}^s = 0$  or vice versa), no observations are missing ( $y_{g1,n}^s = y_{g2,n}^s = 1$ ), and two observations are missing ( $y_{g1,n}^s = y_{g2,n}^s = 0$ ). To simplify notation, we therefore define an index set  $\mathcal{A} = \{10, 01, 11, 00\}$  based on the values that  $y_{g1,n}^s$  and  $y_{g2,n}^s$  take and the corresponding indicator functions  $d_{g,n}^a = \mathbb{1}(10y_{g1,n}^s + y_{g2,n}^s = a)$ . In order to construct the likelihood function, we also need to introduce some additional notation. Let  $\zeta_{g,n} = 2y_{g,n}^s - \iota_2$ , where  $\iota_2$  is the 2-dimensional vector of ones,  $\tilde{S}_{g,n}^s(\lambda) = \left( \zeta_{g1,n} S_{g1,n}^{s'}(\lambda), \zeta_{g2,n} S_{g2,n}^{s'}(\lambda) \right)'$ , and

$$\tilde{\Omega}_{g,n}^{ss}(\theta) = \begin{pmatrix} \Omega_{g11,n}^{ss}(\theta) & \zeta_{g1,n} \zeta_{g2,n} \Omega_{g12,n}^{ss}(\theta) \\ \zeta_{g1,n} \zeta_{g2,n} \Omega_{g21,n}^{ss}(\theta) & \Omega_{g22,n}^{ss}(\theta) \end{pmatrix}, \quad \tilde{\Omega}_{g,n}^{so}(\theta) = - \begin{pmatrix} \zeta_{g1,n} \Omega_{g11,n}^{so}(\theta) & \zeta_{g1,n} \Omega_{g12,n}^{so}(\theta) \\ \zeta_{g2,n} \Omega_{g21,n}^{so}(\theta) & \zeta_{g2,n} \Omega_{g22,n}^{so}(\theta) \end{pmatrix}$$

with  $\Omega_n^{ss}(\theta) = S_n^s(\lambda^s) S_n^{s'}(\lambda^s)$ ,  $\Omega_n^{so}(\theta) = S_n^s(\lambda^s) S_n^{o'}(\lambda^o) \rho \sigma$ , and  $\Omega_n^{oo}(\theta) = S_n^o(\lambda^o) S_n^{o'}(\lambda^o) \sigma^2$ . Further, let  $z_{g,n}(\theta) = y_{g,n}^o - S_{g,n}^o(\lambda^o) X_n^o \beta^o$ , and for any  $a \in \mathcal{A}$ ,  $R_{g,n}^a(\theta)$  be the correlation matrix obtained from  $\Sigma_{g,n}^a(\theta)$ ,  $v_{g,n}^a(\theta) = (\text{Diag}(\Sigma_{g,n}^a(\theta)))^{-1/2} (q_{g,n}(\theta) - \mu_{g,n}^a(\theta))$  with  $q_{g,n}(\theta) = \tilde{S}_{g,n}^s(\lambda^s) X_n^s \beta^s$ , and

$$\begin{aligned} \mu_{g,n}^{10}(\theta) &= \tilde{\Omega}_{g1,n}^{so}(\theta) z_{g1,n}(\theta) / \Omega_{g11,n}^{oo}(\theta), & \Sigma_{g,n}^{10}(\theta) &= \tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g1,n}^{so}(\theta) \tilde{\Omega}_{g1,n}^{so'}(\theta) / \Omega_{g11,n}^{oo}(\theta), \\ \mu_{g,n}^{01}(\theta) &= \tilde{\Omega}_{g2,n}^{so}(\theta) z_{g2,n}(\theta) / \Omega_{g22,n}^{oo}(\theta), & \Sigma_{g,n}^{01}(\theta) &= \tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g2,n}^{so}(\theta) \tilde{\Omega}_{g2,n}^{so'}(\theta) / \Omega_{g22,n}^{oo}(\theta), \\ \mu_{g,n}^{11}(\theta) &= \tilde{\Omega}_{g,n}^{so}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta), & \Sigma_{g,n}^{11}(\theta) &= \tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g,n}^{so}(\theta) \Omega_{g,n}^{oo-1}(\theta) \tilde{\Omega}_{g,n}^{so'}(\theta), \\ \mu_{g,n}^{00}(\theta) &= 0, & \Sigma_{g,n}^{00}(\theta) &= \tilde{\Omega}_{g,n}^{ss}(\theta). \end{aligned}$$

Then the log-likelihood function based on a grouping  $\mathcal{G}_n$  is defined by (see the derivation in Appendix

C.1)

$$\begin{aligned}
Q_n(\theta) &= \frac{1}{n} \sum_{g \in \mathcal{G}_n} \left\{ \mathbb{1}(y_{g1,n}^s = 1, y_{g2,n}^s = 0) \ln \left[ \frac{1}{\sqrt{\Omega_{g11,n}^{oo}(\theta)}} \phi \left( \frac{z_{g1,n}(\theta)}{\sqrt{\Omega_{g11,n}^{oo}(\theta)}} \right) \Phi_2(v_{g,n}^{10}(\theta), R_{g,n}^{10}(\theta)) \right] \right. \\
&\quad + \mathbb{1}(y_{g1,n}^s = 0, y_{g2,n}^s = 1) \ln \left[ \frac{1}{\sqrt{\Omega_{g22,n}^{oo}(\theta)}} \phi \left( \frac{z_{g2,n}(\theta)}{\sqrt{\Omega_{g22,n}^{oo}(\theta)}} \right) \Phi_2(v_{g,n}^{01}(\theta), R_{g,n}^{01}(\theta)) \right] \\
&\quad + \mathbb{1}(y_{g1,n}^s = 1, y_{g2,n}^s = 1) \ln [\phi_2(z_{g,n}(\theta), \Omega_{g,n}^{oo}(\theta)) \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))] \\
&\quad \left. + \mathbb{1}(y_{g1,n}^s = 0, y_{g2,n}^s = 0) \ln [\Phi_2(v_{g,n}^{00}(\theta), R_{g,n}^{00}(\theta))] \right\} \\
&= \frac{1}{n} \sum_{g \in \mathcal{G}_n} \sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta),
\end{aligned} \tag{4}$$

where  $f_{g,n}^a(\theta)$ ,  $a \in \mathcal{A}$ , represent the log-density functions,  $\phi(\cdot)$  is the standard normal density function, and  $\phi_2(\cdot, \Sigma)$  and  $\Phi_2(\cdot, \Sigma)$  are the bivariate normal density and distribution functions, respectively, with zero mean and variance matrix  $\Sigma$ .

Although the log-likelihood function looks complicated, it is not difficult to implement and to maximize. If there is the spatial error instead of the spatial lag in the selection or outcome equations,  $z_{g,n}(\theta)$  and  $q_{g,n}(\theta)$  have to be replaced with  $z_{g,n}^e(\theta) = y_{g,n}^o - X_{g,n}^o \beta^o$  and  $q_{g,n}^e(\theta) = \tilde{X}_{g,n}^s \beta^s$ , respectively, where  $\tilde{X}_{g,n}^s$  is constructed in the same way as  $\tilde{S}_{g,n}^s(\lambda)$ .

## 4 Asymptotic Properties of Partial Maximum Likelihood Estimator

The main difficulty in proving asymptotic properties of PMLE stems from analyzing the nonlinear objective function based on heterogeneous and spatially dependent processes. Hence, this dependence has to be restricted to a manageable degree. We do so by employing the near epoch dependent (NED) random fields framework developed by Jenish and Prucha (2012). We consider a topological structure proposed in their paper. Let the location of an observation  $i$  be defined by  $l_i \in \tilde{\mathfrak{D}}_n$ , where  $\tilde{\mathfrak{D}}_n$  is a finite sample region of a  $\tilde{d}$ -dimensional lattice  $\tilde{\mathfrak{D}} \subset \mathbb{R}^{\tilde{d}}$ ,  $\tilde{d} > 1$ , equipped with the Euclidean metric. Since the likelihood function in (4) is in terms of likelihood contributions for pairs, let a group  $g = (i, j)'$  be assigned a location  $l_g = (l'_{g1}, l'_{g2})' = (l'_i, l'_j)' \in \tilde{\mathfrak{D}}_n \times \tilde{\mathfrak{D}}_n = \mathfrak{D}_n$ , which is a finite sample region of a  $2\tilde{d}$ -dimensional lattice  $\mathfrak{D} = \tilde{\mathfrak{D}} \times \tilde{\mathfrak{D}} \subset \mathbb{R}^{2\tilde{d}}$ . Given this definition, the distance between two groups  $g$  and  $\dot{g}$  depends on configurations of four points

in  $\mathbb{R}^{\bar{d}}$ . Similarly to Bai et al. (2012), we consider a distance metric between two points in  $\mathbb{R}^{2\bar{d}}$  defined by  $d(l_g, l_{\dot{g}}) = \min\{\|l_{g_1} - l_{\dot{g}_1}\|, \|l_{g_1} - l_{\dot{g}_2}\|, \|l_{g_2} - l_{\dot{g}_1}\|, \|l_{g_2} - l_{\dot{g}_2}\|\}$ , i.e. the minimum distance between two points in sets  $(l_{g_1}, l_{g_2})$  and  $(l_{\dot{g}_1}, l_{\dot{g}_2})$ . The distance between any two subsets  $\mathfrak{A}, \mathfrak{B} \subseteq \mathfrak{D}$  is defined as  $d(\mathfrak{A}, \mathfrak{B}) = \inf\{d(g, \dot{g}) : l_g \in \mathfrak{A}, l_{\dot{g}} \in \mathfrak{B}\}$ , where the fact that the observations are indexed by natural numbers allows us to write  $d(g, \dot{g}) \equiv d(l_g, l_{\dot{g}})$  for two groups  $g$  and  $\dot{g}$  with locations  $l_g, l_{\dot{g}} \in \mathbb{R}^{2\bar{d}}$ .

**Assumption 3.** *Individual units in the economy are located or living in a region  $\tilde{\mathfrak{D}}_n \subset \tilde{\mathfrak{D}} \subset \mathbb{R}^{\bar{d}}$ . The cardinality of  $\mathfrak{D}_n = \tilde{\mathfrak{D}}_n \times \tilde{\mathfrak{D}}_n$  satisfies  $\lim_{n \rightarrow \infty} |\mathfrak{D}_n| = \infty$ . The distance  $d(g, \dot{g})$  between any two different groups  $g$  and  $\dot{g}$  is larger than or equal to a specific positive constant, which we normalize to 1.*

Region  $\mathfrak{D}$  corresponds to a space of economic or geographic characteristics or a mixture of them. In Example 2, a geographical space can simply be used. Although there is no natural location for an observation  $i$  in Example 1, a location can be constructed. Assume that there are  $t = 1, \dots, T$  tutorial groups with at most  $\bar{S}$  students in each group. Let  $t_i$  be the tutorial group of student  $i$  and  $a_i$  be his rank in tutorial group  $t_i$  based on the alphabetical ordering. Then student  $i$ 's location can be given by  $l_i = (t_i \bar{S}, a_i)'$ . Assumption 3 implies that the increasing domain asymptotics is used (as an alternative to the infill domain asymptotics): the distance restriction in Assumption 3 implies that there is a finite number of units in any bounded region and that the sample region  $\mathfrak{D}_n$  has to expand when the sample grows.

For reference, the definitions of  $\alpha$ -mixing and NED properties presented in Jenish and Prucha (2009, 2012) are reviewed first.

**Definition 1.** *Let  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  be a triangular array of real random variables defined on a probability space  $(\Omega, \mathcal{F}, P)$ . Moreover, let  $\mathcal{A}$  and  $\mathcal{B}$  be two  $\sigma$ -algebras of  $\mathcal{F}$  and*

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup(|P(A \cap B) - P(A)P(B)|, A \in \mathcal{A}, B \in \mathcal{B}).$$

*For  $\mathfrak{A} \subseteq \mathfrak{D}_n$  and  $\mathfrak{B} \subseteq \mathfrak{D}_n$ , let  $\sigma_n(\mathfrak{A}) = \sigma(\eta_{g,n} : l_g \in \mathfrak{A})$  and  $\alpha_n(\mathfrak{A}, \mathfrak{B}) = \alpha(\sigma_n(\mathfrak{A}), \sigma_n(\mathfrak{B}))$ . Then the  $\alpha$ -mixing coefficients for the random fields  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  are defined as*

$$\bar{\alpha}(k, m, s) = \sup_n \sup_{\mathfrak{A}, \mathfrak{B}} (\alpha_n(\mathfrak{A}, \mathfrak{B}), |\mathfrak{A}| \leq k, |\mathfrak{B}| \leq m, d(\mathfrak{A}, \mathfrak{B}) \geq s),$$

*where  $|\cdot|$  denotes the cardinality of a set.*

This definition is similar to the time series literature. The major difference is that, in the random fields setting, the  $\alpha$ -mixing coefficients do not only depend on the distance between two sets but also on the sizes

of the sets. The definition of the NED property follows.

**Definition 2.** Let  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  and  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  be random fields located on  $\mathfrak{D}_n$ , and additionally,  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  satisfy  $\|Z_{g,n}\|_p < \infty$ ,  $p \geq 1$ . Moreover, let  $\{t_{g,n}\}_{g \in \mathcal{G}_n}$  be an array of positive constants. Then the random field  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  is said to be  $L_p$ -near epoch dependent on the random field  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  if

$$\|Z_{g,n} - E[Z_{g,n} | \mathcal{F}_{g,n}(s)]\|_p \leq t_{g,n} \psi(s)$$

for some sequence  $\psi(s) \geq 0$  with  $\lim_{s \rightarrow \infty} \psi(s) = 0$ , where  $\mathcal{F}_{g,n}(s) = \sigma(\eta_{\dot{g},n} : d(g, \dot{g}) \leq s)$ . The NED random field is uniform if and only if  $\sup_{n,g} t_{g,n} < \infty$ .

In Definition 2, the term  $Z_{g,n} - E[Z_{g,n} | \mathcal{F}_{g,n}(s)]$  measures the prediction error of  $Z_{g,n}$  based on the information contained in  $\{\eta_{\dot{g},n} : d(g, \dot{g}) \leq s\}$ . The NED property then states that the prediction error converges to zero as  $s$  increases. Note that NED is not a property of a random variable itself as  $\alpha$ -mixing is, but it is a property of a mapping.

## 4.1 Consistency

To prove consistency, we need to introduce additional assumptions.

**Assumption 4.** (i)  $\{(X_{g,n}^s, X_{g,n}^o)\}_{g \in \mathcal{G}_n}$  is an  $\alpha$ -mixing random field with  $\alpha$ -mixing coefficients  $\bar{\alpha}(k, m, s) \leq (k + m)^\tau \hat{\alpha}(s)$ ,  $\tau \geq 0$ , for some  $\hat{\alpha}(s) \rightarrow 0$  as  $s \rightarrow \infty$  such that  $\sum_{s=1}^{\infty} s^{2\tilde{d}-1} \hat{\alpha}(s) < \infty$ . (ii)  $\sup_{n,i} E\|X_{i,n}^b\|^p < \infty$  and  $\sup_{n,i} E[\|X_{i,n}^b\|^p | y_{i,n}^s = 1] < \infty$  for any given  $p \geq 1$ , where  $b \in \{s, o\}$ .

Assumption 4(i) states that the exogenous variables may be cross-sectionally dependent under some restrictions. Assumption 4(ii) implies that infinitely many moments of the exogenous variables exist. This is less restrictive than the assumption that the support of the exogenous variables is bounded, which is usual in the (spatial) literature studying discrete choice or limited dependent variable models.<sup>10</sup> (Note that the proofs require only finitely many moments but with a quite large  $p$  and that finding the exact  $p$  would require a lot of effort such as calculating the third order derivatives of the bivariate normal distribution functions in (4).) Moreover, the large number  $p$  of finite moments is related to the assumed normality of the errors and the need to bound the moments of the logarithm of the bivariate normal cumulative distribution function and their derivatives. For heavier-tailed error distributions (e.g., the Laplace distribution), a substantially smaller number of moments would have to exist.

<sup>10</sup>E.g., see condition (v) of Theorem 1 by Pinkse and Slade (1998) and condition (vi) of Theorem 1 by Wang et al. (2013).

The elements of the spatial weight matrices determine the strength of dependence between observations. An important question is under which structures of spatial weight matrices the limit laws based on the NED framework hold. Given that the likelihood function is specified in terms of (the inverses  $S_n^s(\lambda^s)$  and  $S_n^o(\lambda^o)$  of)  $I_{2n} - \lambda^s W_n^s$  and  $I_{2n} - \lambda^o W_n^o$  and the grouping is determined by the user of the method, we impose restrictions on the weight matrices indirectly by the following assumption.

**Assumption 5.**  $\lim_{s \rightarrow \infty} \psi(s) = 0$  with  $\psi(s) = \max\{\psi^s(s), \psi^o(s)\}$ , where  $\psi^b(s) = \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g}: d(g,\dot{g}) > s} \|S_{g\dot{g},n}^b(\lambda^b)\| / \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda^b)\|$ ,  $b \in \{s, o\}$ .

As shown in the proof of Theorem 1, Assumption 5 is needed to show that  $\{\sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta)\}_{g \in \mathcal{G}_n}$  is a NED random field on the  $\alpha$ -mixing random field  $\{\eta_{g,n} = (X_{g,n}^s, X_{g,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g \in \mathcal{G}_n}$ . Although it is not clear how to find the conditions for the weight matrices and groupings that imply Assumption 5, it is possible to check whether weight matrices with certain structures and proposed grouping schemes satisfy this assumption. For instance, if tutorial groups are analyzed in Example 1, it is typically assumed that the  $ij$ th element of the weight matrix is equal to zero if students  $i$  and  $j$  are from different tutorial groups. Thus assuming that the number of students in each tutorial group is even, it is beneficial to form only pairs of students who are in the same tutorial group. Given the definition of locations for Example 1 presented below Assumption 3, if  $d(g, \dot{g}) \geq \bar{S}$ , students in pairs  $g$  and  $\dot{g}$  are from different tutorial groups implying that  $\|S_{g\dot{g},n}^b(\lambda^b)\| = 0$  and Assumption 5 is trivially satisfied.

Next, we make an assumption about the  $2 \times 2$  submatrices of matrices  $\Omega_n^{ss}(\theta)$  and  $\Omega_n^{oo}(\theta)$  and  $\Sigma_{g,n}^a(\theta)$ ,  $a \in \mathcal{A}$ , defined in Section 3.

**Assumption 6.** *The minimum eigenvalues of matrices  $\Omega_{g,n}^{ss}(\theta)$ ,  $\Omega_{g,n}^{oo}(\theta)$ , and  $\Sigma_{g,n}^a(\theta)$ ,  $a \in \mathcal{A}$ , are bounded away from zero uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$ , where  $\Theta$  is the parameter space of  $\theta$ .*

Assumption 6 ensures that the above mentioned  $2 \times 2$  (sub)matrices are invertible for each pair. Thus, the observations should be grouped in such a way that this assumption is not violated. Its validity can be checked using a grid covering possible values of the spatial parameters and the correlation coefficient since matrices  $\Omega_{g,n}^{ss}(\theta)$ ,  $\Omega_{g,n}^{oo}(\theta)/\sigma^2$ , and  $\Sigma_{g,n}^a(\theta)$ ,  $a \in \mathcal{A}$ , do not depend on regression parameters  $\beta^s$ ,  $\beta^o$ , and variance  $\sigma^2$ .

**Assumption 7.** *The parameter space  $\Theta$  is a compact subset of  $\mathbb{R}^L$ .*

**Assumption 8.** *The population log-likelihood function  $Q_0(\theta)$  is uniquely maximized at  $\theta_0$ .*

Whereas Assumption 7 is a standard assumption for nonlinear extremum estimators, Assumption 8 is the identification condition for PMLE. It is often possible to verify identification in finite samples, that is, for



given weight matrices, but it is very difficult to establish more primitive conditions for the limiting objective function in Assumption 8 due to possibly degenerate limits of  $W_n^s$  and  $W_n^o$  and the corresponding possibly “unbounded” heterogeneity and dependence of observations (see a discussion in Xu and Lee, 2015b, footnote 5).

Finally, the consistency of the proposed PMLE follows.

**Theorem 1.** *Under Assumptions 1–8,  $\hat{\theta}_n - \theta_0 = o_p(1)$  as  $n \rightarrow \infty$ .*

## 4.2 Asymptotic Normality

In order to establish asymptotic normality, we need to strengthen the assumptions regarding the dependence structure. In particular, the NED coefficients of random field  $\{\sum_{a \in \mathcal{A}} d_{g,n}^a \partial f_{g,n}^a(\theta_0)/\partial \theta\}_{g \in \mathcal{G}_n}$  have to decrease to zero at a certain relatively fast rate, see Assumption 10 below, because the likelihood function is not only nonlinear but contains indicator functions as well. Additionally, some standard regularity conditions are required.

**Assumption 9.**  $\{(X_{g,n}^s, X_{g,n}^o)\}_{g \in \mathcal{G}_n}$  is an  $\alpha$ -mixing random field with  $\alpha$ -mixing coefficients  $\bar{\alpha}(k, m, s) \leq (k + m)^\tau \hat{\alpha}(s)$  with some  $\hat{\alpha}(s) \rightarrow 0$  as  $s \rightarrow \infty$  such that for some  $\delta > 0$ ,  $\sum_{s=1}^{\infty} s^{2\bar{d}(\tau_*+1)-1} \hat{\alpha}^{\delta/(4+2\delta)}(s) < \infty$ , where  $\tau_* = \delta\tau/(2 + \delta)$ ,  $\tau \geq 0$ .

**Assumption 10.** The NED coefficients satisfy  $\sum_{s=1}^{\infty} s^{2\bar{d}-1} \psi^{(r-2)/(12r-12)}(s) < \infty$ , for some  $r > 2$  with  $\psi(s)$  defined in Assumption 5.

**Assumption 11.**  $\theta_0$  is in the interior of the parameter space  $\Theta$ .

**Assumption 12.** (i)  $H(\theta_0) = \lim_{n \rightarrow \infty} E \left[ \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} \right]$  exists, is finite and nonsingular. (ii) The minimum eigenvalue of  $J_n(\theta_0) = E \left[ n \frac{\partial Q_n(\theta_0)}{\partial \theta} \frac{\partial Q_n(\theta_0)}{\partial \theta'} \right]$  is bounded away from zero uniformly in  $n \in \mathbb{N}$ . (iii)  $J(\theta_0) = \lim_{n \rightarrow \infty} J_n(\theta_0)$  exists and is finite.

Given the non-singular Jacobian and Hessian matrices corresponding to the population partial maximum likelihood function, the asymptotic normality follows.

**Theorem 2.** *Under Assumptions 1–3, 4(ii)–12,  $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, H^{-1}(\theta_0)J(\theta_0)H^{-1}(\theta_0))$  as  $n \rightarrow \infty$ .*

Finally, since the likelihood function does not account for the dependence between groups, the variance matrix of PMLE is not equal to the inverse of the Fisher information matrix. Thus, PMLE is in general not efficient as the full MLE is.

### 4.3 Estimation of the Variance Matrix

Although we do not model the dependence between pairs in the likelihood function, it has to be accounted for when the variance matrix is estimated. On the one hand, it is relatively easy to estimate the Hessian matrix  $H_n(\theta_0) = E[\partial^2 Q_n(\theta_0)/\partial\theta\partial\theta']$  as it can be obtained using its sample analog and a consistent estimate of  $\theta$ :  $\hat{H}_n(\hat{\theta}_n) = \partial^2 Q_n(\hat{\theta}_n)/\partial\theta\partial\theta'$ . On the other hand, estimation of the variance matrix  $J_n(\theta_0)$  and its limit  $J(\theta_0)$  is complicated due to the dependence between groups. It is theoretically possible to consider a spatial analog of a heteroskedasticity and autocorrelation consistent (HAC) estimator of the variance matrix that has been extensively analyzed in the time series literature (i.e., Newey and West, 1987, and Andrews, 1991). Conley (1999) adapted the HAC estimator for the spatially stationary observations. Noting that the Cliff-Ord type models are in general not spatially stationary, Kelejian and Prucha (2007) and Kim and Sun (2011) relaxed the stationarity assumption, but considered only processes linear in error terms. This is not the case for  $\partial Q_n(\theta_0)/\partial\theta$  here, and therefore, the HAC estimator is not easily applicable in the present setting.

On the other hand, it is not uncommon to estimate the variance of an estimator of a spatial model using the bootstrap when it is very difficult or practically impossible to obtain a closed form expression of the variance matrix (e.g., a residual based bootstrap method is used by Su and Yang (2015) in spatial dynamic panel data models). Given that the considered sample selection models are completely parametrically specified, it is possible to use the parametric bootstrap to estimate  $J_n(\theta_0)$ . Note that we suggest to bootstrap  $J_n(\theta_0)$  instead of the complete variance matrix of the estimator to guarantee good computational speed. The bootstrap procedure for estimating  $J_n(\theta_0)$  can be described for the spatial lag model as follows (the spatial error model can be dealt with analogously).

1. Obtain the partial maximum likelihood estimate  $\hat{\theta}_n = (\hat{\beta}_n^{s'}, \hat{\beta}_n^{o'}, \hat{\lambda}_n^s, \hat{\lambda}_n^o, \hat{\rho}_n, \hat{\sigma}_n^2)'$ .
2. For every  $b = 1, \dots, B$ , generate a random sample  $(u_{i,n}^{s(b)}, u_{i,n}^{o(b)})'$  of size  $2n$  from the distribution  $\mathcal{N}(0, \Sigma(\hat{\theta}_n))$ , where  $\Sigma(\hat{\theta}_n) = [1 \ \hat{\rho}_n \hat{\sigma}_n; \ \hat{\rho}_n \hat{\sigma}_n \ \hat{\sigma}_n^2]$ .
3. Given  $W_n^s, W_n^o, X_n^s, X_n^o$ , and  $\hat{\theta}_n$ , generate the bootstrap data (indexed by  $b$ ) according to

$$y_n^{*s(b)} = S_n^s(\hat{\lambda}_n^s) X_n^s \hat{\beta}_n^s + \varepsilon_n^{s(b)}(\hat{\lambda}_n^s)$$

$$y_n^{*o(b)} = S_n^o(\hat{\lambda}_n^o) X_n^o \hat{\beta}_n^o + \varepsilon_n^{o(b)}(\hat{\lambda}_n^o)$$

with observed responses  $y_{i,n}^{s(b)} = \mathbf{1}(y_{i,n}^{*s(b)} > 0)$  and  $y_{i,n}^{o(b)} = y_{i,n}^{s(b)} y_{i,n}^{*o(b)}$ , where  $\varepsilon_n^{s(b)}(\hat{\lambda}_n^s) = S_n^s(\hat{\lambda}_n^s) u_n^{s(b)}$  and  $\varepsilon_n^{o(b)}(\hat{\lambda}_n^o) = S_n^o(\hat{\lambda}_n^o) u_n^{o(b)}$ .

4. Compute the score  $\Gamma_n^{(b)}(\hat{\theta}_n) = \partial Q_n^{(b)}(\hat{\theta}_n)/\partial\theta$  for  $B$  bootstrap samples;  $Q_n^{(b)}(\hat{\theta}_n)$  is thus obtained using

$y_{i,n}^{s(b)}, y_{i,n}^{o(b)}, W_n^s, W_n^o, X_n^s$ , and  $X_n^o$ , where  $b = 1, \dots, B$ .

5. Finally, the bootstrap estimate of  $J_n(\theta_0)$  is given by

$$\hat{J}_n(\hat{\theta}_n) = \frac{n}{B-1} \sum_{b=1}^B \left( \Gamma_n^{(b)}(\hat{\theta}_n) - \frac{1}{B} \sum_{b=1}^B \Gamma_n^{(b)}(\hat{\theta}_n) \right) \left( \Gamma_n^{(b)}(\hat{\theta}_n) - \frac{1}{B} \sum_{b=1}^B \Gamma_n^{(b)}(\hat{\theta}_n) \right)'.$$

Since the functional form of derivatives of the likelihood function is rather complicated, we suggest to use numerical differentiation to evaluate  $\Gamma_n^{(b)}$  in step 4. As our simulation study shows (see Section 5), numerical differentiation in this setting works well.

## 5 Monte Carlo Simulations

### 5.1 Experimental Design

We consider the following data generating process:

$$y_n^{*s} = (I_{2n} - \lambda^s W_n^s)^{-1} (X_n^s \beta^s + u_n^s)$$

$$y_n^{*o} = (I_{2n} - \lambda^o W_n^o)^{-1} (X_n^o \beta^o + u_n^o)$$

for the spatial lag model and

$$y_n^{*s} = X_n^s \beta^s + (I_{2n} - \lambda^s W_n^s)^{-1} u_n^s$$

$$y_n^{*o} = X_n^o \beta^o + (I_{2n} - \lambda^o W_n^o)^{-1} u_n^o$$

for the spatial error model, where  $X_{i,n}^s = (X_{i1,n}^s, X_{i2,n}^s, X_{i3,n}^s)$  and  $X_{i,n}^o = (X_{i1,n}^o, X_{i2,n}^o, X_{i3,n}^o)$  with  $X_{i1,n}^s = X_{i1,n}^o = 1$ ,  $X_{i2,n}^s = X_{i2,n}^o \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ ,  $X_{i3,n}^s \stackrel{iid}{\sim} \chi^2(1)$ , and  $X_{i3,n}^o \stackrel{iid}{\sim} \chi^2(1)$ . The error terms  $(u_{i,n}^s, u_{i,n}^o)' \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma)$ , where  $\Sigma = [1 \ 0.5; 0.5 \ 1]$ . The parameters  $(\beta_2^s, \beta_3^s, \beta_1^o, \beta_2^o, \beta_3^o) = (1, -1, 1, 1, -1)$ , while  $\beta_1^s$  is chosen for each Monte Carlo iteration in such a way that  $y_{i,n}^s = 0$  for one third of the observations. We analyze all possible combinations of spatial parameters  $\lambda^s$  and  $\lambda^o$  taking values from the set  $\{0, 0.4, 0.85\}$ .

Let matrix  $D_0$  represent the great-cycle distances in miles between 3219 counties in the US in 2000.<sup>11</sup> For each Monte Carlo iteration, we draw a uniform random natural number  $r$  between 1 and  $3220 - 2n$  and construct a distance matrix  $D_n$  as the  $2n \times 2n$  submatrix of  $D_0$  with its upper left corner at  $(r, r)$ . The weight matrices are generated as follows:  $\widetilde{W}_{ij,n}^s = \mathbb{1}(D_{ij,n} \leq 50) \cdot 1/D_{ij,n}^2$  and  $\widetilde{W}_{ij,n}^o = \mathbb{1}(D_{ij,n} \leq 50) \cdot 1/D_{ij,n}$ .

<sup>11</sup>The data is available at <http://data.nber.org/data/county-distance-database.html>.

We row-normalize  $\widetilde{W}_n^s$  and  $\widetilde{W}_n^o$  in order to get  $W_n^s$  and  $W_n^o$ . Wang et al. (2013) propose to group adjacent observations, for instance, based on the Euclidean distance. Based on this idea, we formulate the following integer linear programming (ILP) problem:<sup>12</sup>

$$\begin{aligned} & \min_{e_{ij,n}: i=1,\dots,2n, j=1,\dots,2n} \sum_{i=1}^{2n} \sum_{j=1}^{2n} e_{ij,n} \widetilde{D}_{ij,n} \\ \text{s.t. } & \sum_{j=1}^{2n} e_{ij,n} = 1, \quad e_{ii,n} = 0, \quad e_{ij,n} = e_{ji,n}, \end{aligned}$$

where  $e_{ij,n}$  is a binary variable, which is equal to 1 if observations  $i$  and  $j$  form a pair and to zero otherwise;  $\widetilde{D}_{ij,n} = \mathbb{1}(D_{ij,n} \leq 50)D_{ij,n}$ . We use  $\widetilde{D}_n$  instead of  $D_n$  in the ILP problem in order to reduce the burden of computation. The effect on the grouping is small since the algorithm groups nearby observations.

The PMLE estimator is compared to the ordinary MLE estimator and Heckman's two step estimator (HE), which do not take into account spatial dependence, and to the HMLE estimator, which has a likelihood function of a form similar to (4) but for univariate rather than bivariate observations. The model with the spatial error is also estimated by the GMM estimator proposed by Flores-Lagunes and Schnier (2012). Two versions of the GMM estimator are explored: with the identity weight matrix (GMM) and with the optimal weight matrix (GMM2).

Let  $h = 1, \dots, H$  denote Monte Carlo iterations. Then the bias and the root mean squared error (RMSE) in Tables 1, 2, and 4 (Appendix A) of  $\beta^s = (\beta_1^s, \beta_2^s, \beta_3^s)'$  and  $\beta^o = (\beta_1^o, \beta_2^o, \beta_3^o)'$  are obtained by  $Bias = \|H^{-1} \sum_{h=1}^H (\hat{\beta}_n^{b(h)} - \beta_0^{b(h)})\|$  and  $RMSE = \left( H^{-1} \sum_{h=1}^H \|\hat{\beta}_n^{b(h)} - \beta_0^{b(h)}\|^2 \right)^{1/2}$ , where  $b \in \{s, o\}$  and  $\hat{\beta}_n^{b(h)}$  and  $\beta_0^{b(h)}$  are the estimates and the true parameters, respectively, of the  $h$ th iteration.<sup>13</sup> For the scalar parameters, the bias (instead of the absolute bias) and RMSE are reported.

In many empirical applications, marginal effects play a crucial role. For this reason, we also consider the following marginal effects:  $mf x_1 = \partial P(y_{i,n}^s = 1 | X_n^s) / \partial X_{j2,n}^s$ ,  $mf x_2 = \partial P(y_{i,n}^s = 1 | X_n^s) / \partial X_{j3,n}^s$ ,  $mf x_3 = \partial E[y_{i,n}^{*o} | y_{i,n}^s = 1, X_n^s, X_n^o] / \partial X_{j2,n}^o$ , and  $mf x_4 = \partial E[y_{i,n}^{*o} | y_{i,n}^s = 1, X_n^s, X_n^o] / \partial X_{j3,n}^o$ ; the formulas are presented in Appendix C.2. (In the spatial error model, the marginal effects are conditional on  $X_{i,n}^s$  and  $X_{i,n}^o$  instead of  $X_n^s$  and  $X_n^o$ .) For spatial lag models, three types of marginal effects might be considered – total, direct, and indirect – as discussed by LeSage and Pace (2009). In this paper, we discuss only total marginal effects; other results are available upon request. Since the marginal effects are different for each

<sup>12</sup>We solve this problem in Matlab using the IBM ILOG CPLEX optimizer, which is free of charge for academics.

<sup>13</sup>The  $\beta_0^{s(h)}$  is iteration specific because the intercept changes for each simulated data set as discussed before.

individual, we use average marginal effects in order to calculate the percentage bias and RMSE, that is,

$$p.bias = 100 \cdot H^{-1} \sum_{h=1}^H \left( \left( \overline{mfx^{(h)}}(\hat{\theta}_n^{(h)}) - \overline{mfx^{(h)}}(\theta_0^{(h)}) \right) / \overline{mfx^{(h)}}(\theta_0^{(h)}) \right),$$

$$RMSE = \left( H^{-1} \sum_{h=1}^H \left( \overline{mfx^{(h)}}(\hat{\theta}_n^{(h)}) - \overline{mfx^{(h)}}(\theta_0^{(h)}) \right)^2 \right)^{1/2},$$

where  $\overline{mfx^{(h)}}(\theta) = (2n)^{-1} \sum_{i=1}^{2n} mfx^{(h)}(i, \theta)$  with  $mfx^{(h)}(i, \theta)$  being one of the four marginal effects evaluated for an individual  $i$  using parameter  $\theta$  at iteration  $h = 1, \dots, H$ .

In order to investigate the finite sample performance of the standard error estimates obtained by the parametric bootstrap defined in Section 4.3, we compare these estimates with the adjusted mean absolute deviation (AMAD), where  $AMAD = \sqrt{\pi/2}MAD$  with  $MAD$  being the mean absolute deviation around the mean obtained from the Monte Carlo experiments (if  $Z$  is normally distributed,  $\sqrt{\pi/2}E|Z - E[Z]|$  is equal to the standard deviation of  $Z$ ). AMAD is considered instead of the standard deviation because in rare cases the estimates of the spatial parameters or the correlation coefficient are very close to the lower bound of the parameter space. Such a rare event related to a small sample size or numerical issues is not relevant for judging the precision of the asymptotic variance computation and estimation. Therefore, we consider a measure which is less sensitive.

Finally, note that the empirical means and root mean squared errors are based on 1000 replications of each experiment. For bootstrapping standard errors, the number of bootstrap samples is chosen to be  $B = 100$  (larger values of  $B$  do not change the results substantially).

## 5.2 Monte Carlo Results

Since the results of MLE and HE are very similar (see Tables 1–5 in Appendix A), only the results for MLE will be discussed.

**Spatial lag model.** First of all, let us discuss the sample selection model with the spatial lag in both the selection and outcome equations. Tables 1 and 2 in Appendix A report biases and RMSEs of the four estimators for sample sizes 150, 300, and 450. In general, these results verify consistency and superiority of PMLE compared to the MLE, HE, and HMLE estimators.

The MLE estimator is biased when the spatial dependence is present; the magnitude of the bias increases with the magnitude of the spatial parameters (see Table 1). This result is predicted by the theory because omitting the spatial lag leads to endogeneity problems. If the spatial dependence is absent, the performance

of HMLE and PMLE is quite similar to MLE. On the other hand, the estimate of  $\lambda^o$  obtained by HMLE is biased if  $\lambda^o = 0.85$  with the average bias being 24%, 15%, and 9% for sample sizes 150, 300, and 450, respectively; the estimate of the correlation coefficient is then slightly biased as well.<sup>14</sup> Further, both PMLE and HMLE have a slight bias for  $\beta^s$ , which gets substantial when  $\lambda^s = 0.85$ . In this case, the bias of HMLE is on average 20%, 10%, and 10% larger when the sample size is 150, 300, and 450, respectively; the magnitude of the bias decreases as the sample size increases. This might be explained by observing that, close to the upper bound of the parameter space of  $\lambda^s$ , small deviations of the spatial parameter in the selection equation have a big impact on the variance of  $\varepsilon_{i,n}^s(\lambda^s)$ . Given that the variance of  $u_{i,n}^s$  is fixed and  $|\partial Q_n(\theta)/\partial \lambda^s| \gg |\partial Q_n(\theta)/\partial \beta^s|$  in a neighborhood of  $\hat{\theta}_n$ , this impact is captured by the estimate of  $\beta^s$ . In terms of the bias in the other parameters, PMLE works very well.

In terms of the RMSE, MLE is in general outperformed by the other two estimators when the spatial dependence is present, whereas PMLE almost always has a smaller RMSE than HMLE (Table 2). The biggest difference arises when  $\lambda^o = 0.85$ . In that case, the estimates of  $\beta^o$  obtained by HMLE have on average 2.4, 2.7, and 2.4 times higher RMSEs when the sample size is 150, 300, and 450, respectively, compared to the estimates obtained by PMLE, whereas the estimates of  $\lambda^o$  obtained by HMLE have on average 5.5, 6.5, and 9 times higher RMSEs than PMLE for  $2n = 150, 300$ , and 450, respectively. Further, as the sample size increases, both the bias and the standard deviation of HMLE and PMLE decrease. This verifies consistency of these estimators. For the rest of our analysis, we consider only the samples with 300 observations.

Regarding the marginal effects, the results obtained by MLE are usually biased when the spatial parameters are not equal to zero (Table 3 in Appendix A); the exceptions are, for example,  $mf x_1$  and  $mf x_2$  that depend only on  $\beta^s$  when  $\lambda^s = 0$  and  $mf x_4$  that depends only on  $\beta^o$  when  $\lambda^o = 0$  as we can see in Appendix C.2. The percentage biases of HMLE and PMLE are relatively small except for  $mf x_3$  when  $\lambda^s = 0.85$  and  $\lambda^o = 0$ . In most of the cases, PMLE performs better than HMLE in terms of the percentage bias of marginal effects, whereas in terms of the RMSE, PMLE performs uniformly better. Specifically when  $\lambda^o = 0.85$ ,  $mf x_3$  and  $mf x_4$  obtained by HMLE have on average 1.7 and 2 times higher RMSEs than PMLE. This result is consistent with our previous findings since  $mf x_3$  and  $mf x_4$  depend on  $\lambda^o$  and  $\beta^o$  and the estimates of these parameters obtained by HMLE have a severe bias and/or large RMSE when  $\lambda^o = 0.85$ , whereas these parameters do not influence  $mf x_1$  and  $mf x_2$ .

**Spatial error model.** Next, we discuss the spatial error sample selection model (Tables 4–5 in Appendix A). There is a large bias in the estimates of  $\beta^s$  obtained by MLE when the spatial error is present in the

<sup>14</sup>In this setting, HMLE finds a local maximum instead of the global maximum. Different starting values for  $\lambda^o$  improve the situation slightly.

selection equation, although this bias does not have an influence on the estimates of  $\beta^o$  and has only a slight influence on the estimates of  $\rho$ . The HMLE estimator in the spatial error case performs much worse compared to the spatial lag case: there are large biases in the estimates of  $\beta^s$ ,  $\lambda^s$ , and  $\lambda^o$  when the spatial parameters in the respective equations are not equal to zero, although severe biases are not present in the estimates of the correlation coefficient. Moreover, we have noticed that HMLE is sensitive to the starting values of the spatial parameters. The estimates of  $\beta^s$  and  $\lambda^s$  obtained by GMM and GMM2 are in most cases severely biased. These results are consistent with the simulation results in Flores-Lagunes and Schnier (2012). Finally, there is a bias in the estimates of  $\beta^s$  obtained by PMLE, which increases with the magnitude of  $\lambda^s$ , whereas the other parameters are estimated well. The estimates of  $\beta^s$  obtained by PMLE exhibit a larger RMSE compared to MLE, whereas the estimates of  $\beta^o$  and  $\rho$  obtained by PMLE have similar or smaller RMSEs.

An interesting feature is here that the bias in the parameter estimates does not have a substantial influence on the marginal effects: for all estimators, the percentage bias is less than 11 percentage points. It might be explained in the following way. For MLE, it is misleading to compare  $\hat{\beta}_n^s$ ,  $\hat{\rho}_n$ , and  $\hat{\sigma}_n^2$  with the true parameters because they estimate different quantities: since the spatial dependence is not accounted for, the MLE estimates adjust to fit the data. Different methods can thus provide similar marginal effects, especially given that we report average marginal effects.<sup>15</sup> Hence, the spatial dependence not accounted for by MLE leads to under- and over-estimation of individual marginal effects, which however average out when evaluating average marginal effects. Although the estimates obtained by the other four estimators can be compared to the true parameters, the estimates adjust to fit the data and biases in the estimates do not necessarily imply that the marginal effects (especially the average marginal effects) are biased. In most cases, the percentage biases obtained by GMM and GMM2 are the largest, whereas RMSEs obtained by GMM and GMM2 compared to the other three estimators are always the largest. The differences among MLE, HMLE, and PMLE in terms of the RMSE arise only for  $mf x_3$  and  $mf x_4$  and only when  $\lambda^o = 0.85$ . In that case, the RMSE of MLE (HMLE) is on average 42% (31%) higher than the RMSE of PMLE.

**Estimates of standard errors.** The performance of the parametric bootstrap is investigated in the context of the spatial lag model. The standard errors (SEs) obtained using the parametric bootstrap are very close to the corresponding AMADs of estimates obtained from the Monte Carlo simulation (Table 6, Appendix A). It confirms that the parametric bootstrap is a valid way to estimate the asymptotic variance matrix and that the asymptotic variance approximates well the finite sample variance of estimates. Moreover, these results show that numerical differentiation works well in this setting.

---

<sup>15</sup>See a similar discussion in Wooldridge (2010, p. 602) for the probit and heteroskedastic probit models.

## 6 Conclusion

This paper examines the sample selection model with a spatial lag of a latent dependent variable or a spatial error in both the selection and outcome equations. We propose to estimate this model by the partial maximum likelihood estimator which is based on the idea that all observations are divided into pairs in such a way that dependence within a pair is more important than dependence between pairs; the likelihood function is constructed as a product of marginal likelihood contributions for these pairs. Since the likelihood function does not capture the dependence between pairs, complexity is reduced and the model can be easily estimated. Using the limit laws for the NED random fields, we establish consistency and asymptotic normality of the PMLE. Our simulation study shows that the proposed estimator performs quite well in small samples, and in most cases, outperforms the ordinary MLE, HE, HMLE, and the GMM estimator proposed by Flores-Lagunes and Schnier (2012). Moreover, PMLE and the developed asymptotic theory can be easily applied to other limited dependent variable models, that is, probit and Tobit models, because the sample selection model has all the components of the former models and they are thus special cases of the sample selection model.

The studied model can be extended in several ways. First, it is based on strong distributional assumptions about the error terms. Although it can be applied to other parametric families of distributions, an interesting exercise is to check the finite sample properties of the quasi-PMLE under non-normality. Second, the asymptotic distribution of the proposed estimator depends on the way how the observations are divided into groups. It is desirable to find an optimal grouping scheme based on some criterion, for example, such that the sum of variances of parameters of interest is minimized. Given the complexity of the variance matrix of PMLE, this is a very difficult task. Nevertheless, as our simulation shows, PMLE performs quite well even with a non-optimal grouping.

## Appendix A Results of the Monte Carlo Experiments



**Table 1:** Biases of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations.

$\lambda^s$	$\lambda^o$		$2n = 150$				$2n = 300$				$2n = 450$			
			MLE	HE	HMLE	PMLE	MLE	HE	HMLE	PMLE	MLE	HE	HMLE	PMLE
0.00	0.00	$\hat{\beta}^s$	0.058	0.063	0.093	0.087	0.033	0.033	0.049	0.047	0.016	0.015	0.027	0.027
		$\hat{\beta}^o$	0.010	0.007	0.012	0.016	0.005	0.002	0.006	0.006	0.002	0.003	0.000	0.001
		$\hat{\lambda}^s$	-	-	-0.014	-0.016	-	-	-0.005	-0.005	-	-	-0.006	-0.007
		$\hat{\lambda}^o$	-	-	-0.002	-0.007	-	-	-0.001	-0.004	-	-	0.001	-0.000
		$\hat{\rho}$	-0.010	-0.015	-0.003	0.001	-0.015	-0.012	-0.012	-0.013	-0.006	-0.004	-0.005	-0.005
		$\hat{\sigma}^2$	0.010	0.009	-0.005	0.002	-0.001	0.005	-0.010	-0.009	-0.002	0.002	-0.009	-0.008
0.00	0.40	$\hat{\beta}^s$	0.081	0.079	0.112	0.109	0.041	0.041	0.055	0.054	0.018	0.016	0.026	0.025
		$\hat{\beta}^o$	0.067	0.063	0.017	0.011	0.064	0.061	0.002	0.002	0.059	0.058	0.003	0.003
		$\hat{\lambda}^s$	-	-	-0.004	-0.007	-	-	0.001	-0.001	-	-	-0.003	-0.004
		$\hat{\lambda}^o$	-	-	-0.011	-0.013	-	-	-0.005	-0.005	-	-	-0.003	-0.004
		$\hat{\rho}$	-0.091	-0.074	-0.039	-0.030	-0.072	-0.059	-0.022	-0.018	-0.063	-0.055	-0.008	-0.005
		$\hat{\sigma}^2$	0.307	0.322	-0.030	-0.020	0.280	0.290	-0.015	-0.011	0.264	0.271	-0.011	-0.007
0.00	0.85	$\hat{\beta}^s$	0.047	0.086	0.103	0.116	0.016	0.030	0.034	0.042	0.024	0.030	0.034	0.038
		$\hat{\beta}^o$	0.803	0.802	0.050	0.029	0.661	0.660	0.047	0.010	0.616	0.615	0.030	0.006
		$\hat{\lambda}^s$	-	-	-0.004	-0.010	-	-	-0.001	-0.001	-	-	0.001	-0.001
		$\hat{\lambda}^o$	-	-	-0.218	-0.023	-	-	-0.145	-0.007	-	-	-0.079	-0.005
		$\hat{\rho}$	-0.298	-0.284	-0.065	0.008	-0.280	-0.274	-0.049	-0.006	-0.277	-0.270	-0.036	0.001
		$\hat{\sigma}^2$	9.617	9.435	0.531	-0.039	8.117	8.056	0.354	-0.056	7.321	7.328	0.213	-0.036
0.40	0.00	$\hat{\beta}^s$	0.131	0.128	0.134	0.130	0.140	0.140	0.071	0.069	0.149	0.148	0.034	0.033
		$\hat{\beta}^o$	0.002	0.001	0.007	0.005	0.005	0.002	0.009	0.007	0.004	0.003	0.005	0.005
		$\hat{\lambda}^s$	-	-	-0.005	-0.009	-	-	0.001	-0.001	-	-	-0.004	-0.005
		$\hat{\lambda}^o$	-	-	-0.010	-0.012	-	-	-0.002	-0.001	-	-	-0.007	-0.004
		$\hat{\rho}$	-0.071	-0.067	-0.006	0.003	-0.058	-0.048	-0.012	-0.009	-0.047	-0.046	0.000	-0.000
		$\hat{\sigma}^2$	0.009	0.014	-0.019	-0.014	0.005	0.012	-0.010	-0.008	-0.001	0.002	-0.010	-0.009
0.40	0.40	$\hat{\beta}^s$	0.134	0.132	0.138	0.128	0.141	0.141	0.067	0.064	0.142	0.143	0.036	0.035
		$\hat{\beta}^o$	0.066	0.066	0.009	0.008	0.068	0.065	0.007	0.007	0.056	0.054	0.004	0.003
		$\hat{\lambda}^s$	-	-	-0.001	-0.005	-	-	0.002	-0.000	-	-	-0.004	-0.005
		$\hat{\lambda}^o$	-	-	-0.009	-0.011	-	-	-0.002	-0.002	-	-	-0.002	-0.004
		$\hat{\rho}$	-0.058	-0.058	-0.020	-0.014	-0.054	-0.045	-0.010	-0.007	-0.042	-0.026	-0.007	-0.003
		$\hat{\sigma}^2$	0.324	0.325	-0.028	-0.019	0.285	0.295	-0.017	-0.013	0.261	0.271	-0.014	-0.011
0.40	0.85	$\hat{\beta}^s$	0.130	0.129	0.117	0.128	0.147	0.144	0.046	0.051	0.146	0.145	0.031	0.038
		$\hat{\beta}^o$	0.793	0.785	0.089	0.039	0.659	0.657	0.058	0.007	0.606	0.607	0.043	0.012
		$\hat{\lambda}^s$	-	-	-0.003	-0.012	-	-	0.007	0.003	-	-	0.004	-0.000
		$\hat{\lambda}^o$	-	-	-0.187	-0.018	-	-	-0.116	-0.011	-	-	-0.090	-0.006
		$\hat{\rho}$	-0.198	-0.177	-0.116	0.013	-0.194	-0.179	-0.083	0.007	-0.170	-0.168	-0.074	-0.008
		$\hat{\sigma}^2$	9.381	9.317	0.441	-0.039	8.190	8.129	0.262	-0.037	7.405	7.407	0.192	-0.021
0.85	0.00	$\hat{\beta}^s$	0.709	0.707	0.415	0.348	0.724	0.722	0.205	0.189	0.719	0.719	0.131	0.119
		$\hat{\beta}^o$	0.005	0.011	0.010	0.012	0.003	0.005	0.005	0.002	0.004	0.008	0.003	0.004
		$\hat{\lambda}^s$	-	-	-0.011	-0.014	-	-	-0.004	-0.006	-	-	-0.002	-0.003
		$\hat{\lambda}^o$	-	-	-0.010	-0.004	-	-	-0.006	-0.004	-	-	-0.004	-0.004
		$\hat{\rho}$	-0.277	-0.267	0.039	0.039	-0.259	-0.247	0.017	0.013	-0.244	-0.236	0.023	0.013
		$\hat{\sigma}^2$	0.050	0.065	-0.037	-0.031	0.031	0.034	-0.013	-0.011	0.019	0.022	-0.011	-0.011
0.85	0.40	$\hat{\beta}^s$	0.710	0.709	0.447	0.370	0.719	0.719	0.194	0.178	0.722	0.723	0.143	0.130
		$\hat{\beta}^o$	0.080	0.072	0.014	0.011	0.047	0.046	0.009	0.007	0.049	0.048	0.004	0.003
		$\hat{\lambda}^s$	-	-	-0.007	-0.012	-	-	-0.006	-0.008	-	-	0.000	-0.001
		$\hat{\lambda}^o$	-	-	-0.008	-0.004	-	-	-0.006	-0.006	-	-	-0.003	-0.004
		$\hat{\rho}$	-0.186	-0.166	0.009	0.016	-0.162	-0.151	-0.003	0.004	-0.161	-0.145	0.000	-0.004
		$\hat{\sigma}^2$	0.383	0.408	-0.046	-0.039	0.298	0.309	-0.029	-0.026	0.280	0.292	-0.012	-0.010
0.85	0.85	$\hat{\beta}^s$	0.748	0.711	0.445	0.367	0.740	0.717	0.204	0.182	0.728	0.714	0.144	0.132
		$\hat{\beta}^o$	0.742	0.733	0.067	0.022	0.628	0.623	0.039	0.018	0.580	0.582	0.034	0.003
		$\hat{\lambda}^s$	-	-	-0.002	-0.007	-	-	-0.001	-0.006	-	-	0.002	-0.001
		$\hat{\lambda}^o$	-	-	-0.203	-0.015	-	-	-0.112	-0.006	-	-	-0.072	-0.004
		$\hat{\rho}$	-0.171	-0.129	-0.187	-0.025	-0.151	-0.124	-0.099	-0.002	-0.112	-0.104	-0.072	-0.002
		$\hat{\sigma}^2$	10.055	9.716	0.406	-0.063	8.395	8.088	0.241	-0.036	7.500	7.375	0.114	-0.042

**Table 2:** RMSEs of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations.

$\lambda^s$	$\lambda^o$		$2n = 150$				$2n = 300$				$2n = 450$			
			MLE	HE	HMLE	PMLE	MLE	HE	HMLE	PMLE	MLE	HE	HMLE	PMLE
0.00	0.00	$\hat{\beta}^s$	0.403	0.401	0.435	0.432	0.253	0.255	0.270	0.269	0.207	0.208	0.216	0.216
		$\hat{\beta}^o$	0.247	0.244	0.256	0.261	0.161	0.172	0.164	0.164	0.134	0.143	0.135	0.135
		$\hat{\lambda}^s$	-	-	0.171	0.157	-	-	0.122	0.115	-	-	0.098	0.093
		$\hat{\lambda}^o$	-	-	0.129	0.122	-	-	0.093	0.087	-	-	0.078	0.072
		$\hat{\rho}$	0.343	0.346	0.363	0.365	0.222	0.251	0.229	0.230	0.170	0.198	0.173	0.173
		$\hat{\sigma}^2$	0.175	0.183	0.176	0.185	0.118	0.130	0.120	0.120	0.096	0.103	0.097	0.096
0.00	0.40	$\hat{\beta}^s$	0.398	0.396	0.434	0.429	0.257	0.257	0.272	0.270	0.205	0.206	0.214	0.213
		$\hat{\beta}^o$	0.318	0.329	0.238	0.230	0.215	0.223	0.152	0.149	0.181	0.188	0.124	0.121
		$\hat{\lambda}^s$	-	-	0.175	0.160	-	-	0.118	0.111	-	-	0.099	0.093
		$\hat{\lambda}^o$	-	-	0.117	0.105	-	-	0.073	0.065	-	-	0.059	0.053
		$\hat{\rho}$	0.361	0.376	0.380	0.363	0.245	0.265	0.241	0.232	0.199	0.217	0.185	0.175
		$\hat{\sigma}^2$	0.396	0.426	0.201	0.186	0.329	0.343	0.131	0.124	0.300	0.311	0.108	0.101
0.00	0.85	$\hat{\beta}^s$	0.425	0.413	0.464	0.446	0.274	0.264	0.287	0.281	0.216	0.213	0.227	0.224
		$\hat{\beta}^o$	1.547	1.478	0.819	0.302	1.139	1.100	0.565	0.183	0.942	0.931	0.443	0.152
		$\hat{\lambda}^s$	-	-	0.185	0.161	-	-	0.131	0.117	-	-	0.100	0.091
		$\hat{\lambda}^o$	-	-	0.597	0.148	-	-	0.492	0.084	-	-	0.359	0.060
		$\hat{\rho}$	0.542	0.470	0.547	0.358	0.420	0.377	0.397	0.239	0.367	0.348	0.305	0.201
		$\hat{\sigma}^2$	11.104	10.820	2.092	0.679	8.841	8.731	1.720	0.290	7.747	7.752	1.390	0.284
0.40	0.00	$\hat{\beta}^s$	0.374	0.373	0.446	0.445	0.276	0.277	0.271	0.269	0.248	0.248	0.217	0.216
		$\hat{\beta}^o$	0.250	0.256	0.242	0.237	0.171	0.181	0.164	0.163	0.131	0.139	0.125	0.124
		$\hat{\lambda}^s$	-	-	0.137	0.125	-	-	0.089	0.084	-	-	0.073	0.068
		$\hat{\lambda}^o$	-	-	0.134	0.122	-	-	0.098	0.090	-	-	0.079	0.069
		$\hat{\rho}$	0.371	0.376	0.363	0.341	0.239	0.258	0.226	0.218	0.183	0.203	0.169	0.164
		$\hat{\sigma}^2$	0.179	0.198	0.171	0.175	0.121	0.131	0.119	0.118	0.091	0.096	0.089	0.088
0.40	0.40	$\hat{\beta}^s$	0.380	0.382	0.436	0.432	0.276	0.278	0.273	0.270	0.239	0.241	0.217	0.216
		$\hat{\beta}^o$	0.339	0.335	0.238	0.227	0.221	0.231	0.154	0.148	0.182	0.194	0.123	0.119
		$\hat{\lambda}^s$	-	-	0.138	0.127	-	-	0.090	0.085	-	-	0.071	0.067
		$\hat{\lambda}^o$	-	-	0.101	0.089	-	-	0.072	0.065	-	-	0.058	0.053
		$\hat{\rho}$	0.365	0.363	0.379	0.355	0.237	0.255	0.225	0.214	0.184	0.208	0.181	0.168
		$\hat{\sigma}^2$	0.423	0.431	0.187	0.182	0.338	0.355	0.133	0.125	0.299	0.314	0.104	0.098
0.40	0.85	$\hat{\beta}^s$	0.395	0.375	0.441	0.432	0.290	0.275	0.271	0.261	0.245	0.241	0.217	0.217
		$\hat{\beta}^o$	1.583	1.510	0.853	0.407	1.128	1.092	0.598	0.202	0.924	0.915	0.403	0.262
		$\hat{\lambda}^s$	-	-	0.144	0.128	-	-	0.101	0.088	-	-	0.076	0.069
		$\hat{\lambda}^o$	-	-	0.554	0.116	-	-	0.440	0.113	-	-	0.390	0.082
		$\hat{\rho}$	0.470	0.404	0.550	0.334	0.373	0.323	0.401	0.225	0.287	0.276	0.331	0.179
		$\hat{\sigma}^2$	10.842	10.754	2.051	0.838	9.032	8.915	1.495	0.419	7.847	7.851	1.283	0.777
0.85	0.00	$\hat{\beta}^s$	0.771	0.769	0.908	0.761	0.751	0.749	0.487	0.459	0.738	0.738	0.366	0.345
		$\hat{\beta}^o$	0.306	0.314	0.214	0.214	0.212	0.211	0.148	0.147	0.168	0.171	0.118	0.116
		$\hat{\lambda}^s$	-	-	0.075	0.067	-	-	0.049	0.044	-	-	0.039	0.034
		$\hat{\lambda}^o$	-	-	0.146	0.125	-	-	0.093	0.084	-	-	0.081	0.074
		$\hat{\rho}$	0.519	0.526	0.420	0.395	0.411	0.403	0.280	0.263	0.357	0.353	0.228	0.205
		$\hat{\sigma}^2$	0.211	0.246	0.153	0.153	0.140	0.148	0.109	0.108	0.105	0.112	0.088	0.086
0.85	0.40	$\hat{\beta}^s$	0.769	0.767	0.947	0.769	0.748	0.749	0.460	0.429	0.741	0.742	0.356	0.335
		$\hat{\beta}^o$	0.382	0.390	0.222	0.211	0.255	0.259	0.146	0.140	0.204	0.214	0.120	0.115
		$\hat{\lambda}^s$	-	-	0.071	0.065	-	-	0.050	0.045	-	-	0.038	0.034
		$\hat{\lambda}^o$	-	-	0.120	0.096	-	-	0.074	0.066	-	-	0.060	0.054
		$\hat{\rho}$	0.469	0.464	0.423	0.385	0.341	0.341	0.271	0.239	0.282	0.284	0.207	0.187
		$\hat{\sigma}^2$	0.496	0.561	0.187	0.162	0.358	0.387	0.126	0.115	0.320	0.339	0.101	0.093
0.85	0.85	$\hat{\beta}^s$	0.808	0.768	0.935	0.822	0.771	0.744	0.489	0.445	0.748	0.734	0.355	0.332
		$\hat{\beta}^o$	1.667	1.534	0.771	0.311	1.252	1.124	0.523	0.240	1.014	0.937	0.387	0.137
		$\hat{\lambda}^s$	-	-	0.073	0.062	-	-	0.047	0.071	-	-	0.036	0.032
		$\hat{\lambda}^o$	-	-	0.576	0.075	-	-	0.428	0.044	-	-	0.347	0.021
		$\hat{\rho}$	0.549	0.438	0.629	0.400	0.447	0.328	0.436	0.255	0.348	0.277	0.342	0.200
		$\hat{\sigma}^2$	12.077	11.574	1.914	0.947	9.398	8.969	1.546	0.685	8.052	7.906	1.051	0.140

**Table 3:** Percentage biases and RMSEs of total marginal effects estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations ( $2n = 300$ ).

$\lambda^s$	$\lambda^o$		MLE		HE		HMLE		PMLE	
			p.bias	RMSE	p.bias	RMSE	p.bias	RMSE	p.bias	RMSE
0.00	0.00	$mf x_1$	0.623	0.017	0.718	0.018	1.345	0.030	1.153	0.029
		$mf x_2$	1.019	0.015	0.940	0.015	1.702	0.029	1.566	0.028
		$mf x_3$	-0.041	0.095	-0.258	0.098	0.359	0.135	0.022	0.131
		$mf x_4$	-0.127	0.050	-0.137	0.050	0.476	0.104	0.133	0.099
0.00	0.40	$mf x_1$	0.842	0.018	0.876	0.018	2.018	0.029	1.633	0.028
		$mf x_2$	0.774	0.016	0.703	0.016	1.959	0.028	1.588	0.026
		$mf x_3$	-42.684	0.615	-42.960	0.620	0.683	0.225	0.316	0.211
		$mf x_4$	-37.020	0.612	-37.021	0.612	0.428	0.196	0.078	0.179
0.00	0.85	$mf x_1$	0.828	0.017	0.855	0.017	2.161	0.032	1.905	0.029
		$mf x_2$	0.119	0.016	0.748	0.015	1.394	0.029	1.664	0.027
		$mf x_3$	-82.533	5.107	-82.688	5.116	-4.831	2.153	0.400	1.176
		$mf x_4$	-78.663	5.100	-78.636	5.098	-5.534	2.168	-0.007	1.084
0.40	0.00	$mf x_1$	-35.873	0.108	-35.867	0.108	2.673	0.046	2.296	0.044
		$mf x_2$	-35.757	0.107	-35.797	0.107	2.328	0.041	1.916	0.040
		$mf x_3$	20.309	0.162	19.992	0.161	-0.319	0.160	-0.366	0.148
		$mf x_4$	0.118	0.051	0.166	0.051	0.720	0.111	0.642	0.103
0.40	0.40	$mf x_1$	-36.228	0.109	-36.279	0.109	2.373	0.048	2.000	0.046
		$mf x_2$	-35.467	0.106	-35.499	0.107	2.838	0.042	2.451	0.040
		$mf x_3$	-36.746	0.486	-36.910	0.489	0.295	0.232	0.009	0.215
		$mf x_4$	-36.896	0.610	-36.877	0.610	1.053	0.197	0.804	0.179
0.40	0.85	$mf x_1$	-36.311	0.110	-36.301	0.109	3.634	0.053	2.377	0.045
		$mf x_2$	-36.540	0.110	-36.020	0.108	2.367	0.045	1.883	0.041
		$mf x_3$	-83.329	4.962	-83.499	4.971	-2.290	2.038	-0.014	1.228
		$mf x_4$	-78.656	5.100	-78.600	5.097	-2.991	2.096	0.086	1.102
0.85	0.00	$mf x_1$	-77.849	0.507	-77.822	0.507	4.750	0.187	3.043	0.166
		$mf x_2$	-77.063	0.502	-77.026	0.501	5.083	0.158	3.488	0.138
		$mf x_3$	113.671	0.490	113.383	0.489	-13.500	0.368	-9.847	0.335
		$mf x_4$	0.138	0.050	0.149	0.050	0.253	0.102	0.299	0.095
0.85	0.40	$mf x_1$	-77.920	0.508	-77.929	0.509	3.685	0.187	2.206	0.165
		$mf x_2$	-76.946	0.502	-76.943	0.502	4.962	0.157	3.421	0.137
		$mf x_3$	-5.214	0.134	-5.353	0.134	-3.066	0.440	-3.019	0.382
		$mf x_4$	-37.229	0.615	-37.215	0.615	0.033	0.206	-0.223	0.187
0.85	0.85	$mf x_1$	-77.578	0.508	-77.596	0.508	5.755	0.202	3.316	0.159
		$mf x_2$	-77.840	0.508	-77.112	0.504	5.779	0.173	3.724	0.138
		$mf x_3$	-79.005	4.105	-79.034	4.106	-1.168	1.944	-1.218	1.199
		$mf x_4$	-78.999	5.122	-78.812	5.110	-3.523	2.025	-0.551	1.008

**Table 4:** Biases and RMSEs of parameter estimates in the context of the sample selection model with a spatial error in both the selection and outcome equations ( $2n = 300$ ).

$\lambda^s$	$\lambda^o$		MLE		HE		HMLE		GMM		GMM2		PMLE	
			bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE	bias	RMSE
0.00	0.00	$\hat{\beta}^s$	0.026	0.258	0.027	0.260	0.028	0.260	0.996	2.494	1.035	2.151	0.105	0.320
		$\hat{\beta}^o$	0.005	0.164	0.005	0.173	0.005	0.164	0.018	0.192	0.026	0.176	0.005	0.167
		$\hat{\lambda}^s$	-	-	-	-	-0.001	0.041	-0.133	0.560	-0.136	0.578	-0.026	0.309
		$\hat{\lambda}^o$	-	-	-	-	0.001	0.059	0.024	0.113	0.022	0.112	-0.016	0.225
		$\hat{\rho}$	-0.011	0.212	-0.014	0.240	-0.010	0.213	0.019	0.345	-0.024	0.277	-0.002	0.220
		$\hat{\sigma}^2$	-0.003	0.117	0.002	0.126	-0.006	0.119	-0.081	0.128	-0.029	0.106	-0.028	0.124
0.00	0.40	$\hat{\beta}^s$	0.042	0.265	0.042	0.267	0.058	0.289	1.032	2.420	1.176	2.300	0.123	0.331
		$\hat{\beta}^o$	0.012	0.189	0.011	0.197	0.012	0.190	0.028	0.214	0.037	0.197	0.007	0.184
		$\hat{\lambda}^s$	-	-	-	-	-0.001	0.118	-0.208	0.618	-0.205	0.647	-0.024	0.309
		$\hat{\lambda}^o$	-	-	-	-	-0.397	0.435	-0.070	0.112	-0.072	0.115	-0.025	0.152
		$\hat{\rho}$	-0.043	0.237	-0.043	0.259	-0.038	0.246	-0.008	0.357	-0.047	0.290	-0.015	0.234
		$\hat{\sigma}^2$	0.118	0.183	0.123	0.190	0.091	0.180	-0.045	0.110	0.006	0.107	-0.019	0.138
0.00	0.85	$\hat{\beta}^s$	0.023	0.263	0.034	0.259	0.229	1.248	1.165	1.987	1.304	2.119	0.109	0.312
		$\hat{\beta}^o$	0.020	0.505	0.007	0.497	0.009	0.477	0.114	0.535	0.094	0.499	0.010	0.425
		$\hat{\lambda}^s$	-	-	-	-	0.007	0.345	-0.325	0.691	-0.305	0.708	-0.032	0.307
		$\hat{\lambda}^o$	-	-	-	-	-0.412	0.701	-0.082	0.099	-0.083	0.098	-0.011	0.041
		$\hat{\rho}$	-0.137	0.296	-0.148	0.294	-0.027	0.331	-0.175	0.373	-0.182	0.346	0.006	0.245
		$\hat{\sigma}^2$	2.728	2.983	2.714	2.961	0.655	1.310	0.383	0.572	0.412	0.555	-0.013	0.197
0.40	0.00	$\hat{\beta}^s$	0.105	0.280	0.106	0.282	0.089	0.295	0.799	2.076	0.953	2.232	0.145	0.898
		$\hat{\beta}^o$	0.004	0.163	0.008	0.172	0.004	0.164	0.011	0.192	0.017	0.172	0.005	0.164
		$\hat{\lambda}^s$	-	-	-	-	-0.406	0.420	-0.549	0.779	-0.541	0.798	-0.040	0.249
		$\hat{\lambda}^o$	-	-	-	-	0.000	0.091	0.000	0.116	-0.004	0.111	-0.012	0.215
		$\hat{\rho}$	-0.031	0.224	-0.023	0.248	-0.029	0.226	0.008	0.363	-0.027	0.289	-0.006	0.230
		$\hat{\sigma}^2$	-0.007	0.121	0.000	0.132	-0.013	0.128	-0.086	0.130	-0.035	0.105	-0.030	0.126
0.40	0.40	$\hat{\beta}^s$	0.119	0.293	0.119	0.294	0.087	0.344	1.012	2.565	1.145	2.752	0.128	0.431
		$\hat{\beta}^o$	0.003	0.193	0.005	0.201	0.003	0.193	0.033	0.226	0.035	0.207	0.004	0.186
		$\hat{\lambda}^s$	-	-	-	-	-0.397	0.424	-0.646	0.892	-0.636	0.902	-0.046	0.274
		$\hat{\lambda}^o$	-	-	-	-	-0.388	0.422	-0.079	0.123	-0.082	0.123	-0.028	0.156
		$\hat{\rho}$	-0.012	0.230	-0.004	0.254	-0.007	0.235	0.004	0.380	-0.024	0.296	0.007	0.218
		$\hat{\sigma}^2$	0.123	0.190	0.131	0.205	0.096	0.186	-0.049	0.122	0.001	0.110	-0.015	0.142
0.40	0.85	$\hat{\beta}^s$	0.117	0.299	0.105	0.287	0.111	0.741	1.114	2.152	1.306	2.534	0.126	0.417
		$\hat{\beta}^o$	0.023	0.506	0.007	0.512	0.019	0.485	0.168	0.564	0.135	0.507	0.006	0.419
		$\hat{\lambda}^s$	-	-	-	-	-0.373	0.546	-0.801	1.004	-0.795	1.012	-0.046	0.266
		$\hat{\lambda}^o$	-	-	-	-	-0.444	0.724	-0.085	0.103	-0.084	0.098	-0.013	0.039
		$\hat{\rho}$	-0.066	0.264	-0.082	0.271	0.040	0.319	-0.144	0.364	-0.149	0.344	0.009	0.221
		$\hat{\sigma}^2$	2.730	2.980	2.717	2.965	0.763	1.399	0.364	0.522	0.394	0.524	0.000	0.175
0.85	0.00	$\hat{\beta}^s$	1.119	1.219	1.119	1.219	0.911	1.578	0.521	1.592	0.292	1.830	0.287	1.407
		$\hat{\beta}^o$	0.006	0.200	0.012	0.207	0.006	0.199	0.014	0.252	0.006	0.208	0.003	0.177
		$\hat{\lambda}^s$	-	-	-	-	-0.831	0.879	-1.004	1.160	-0.974	1.154	-0.014	0.077
		$\hat{\lambda}^o$	-	-	-	-	-0.004	0.117	-0.015	0.117	-0.020	0.117	0.004	0.209
		$\hat{\rho}$	-0.137	0.320	-0.129	0.319	-0.119	0.335	-0.102	0.493	-0.093	0.373	-0.006	0.330
		$\hat{\sigma}^2$	0.026	0.142	0.034	0.169	0.018	0.145	-0.066	0.119	-0.014	0.105	-0.012	0.125
0.85	0.40	$\hat{\beta}^s$	1.100	1.197	1.101	1.198	0.767	1.960	0.302	2.034	0.226	1.824	0.308	1.294
		$\hat{\beta}^o$	0.010	0.220	0.008	0.236	0.008	0.218	0.039	0.297	0.027	0.249	0.008	0.192
		$\hat{\lambda}^s$	-	-	-	-	-0.832	0.897	-1.095	1.265	-1.085	1.273	-0.014	0.079
		$\hat{\lambda}^o$	-	-	-	-	-0.390	0.441	-0.084	0.134	-0.084	0.133	-0.016	0.159
		$\hat{\rho}$	-0.100	0.303	-0.087	0.320	-0.074	0.318	-0.069	0.523	-0.068	0.392	-0.012	0.282
		$\hat{\sigma}^2$	0.148	0.225	0.167	0.261	0.112	0.220	-0.038	0.117	0.012	0.109	-0.014	0.140
0.85	0.85	$\hat{\beta}^s$	1.102	1.202	1.082	1.182	0.074	3.026	0.418	1.494	0.231	1.679	0.349	1.404
		$\hat{\beta}^o$	0.032	0.562	0.012	0.559	0.008	0.505	0.292	0.657	0.260	0.607	0.007	0.412
		$\hat{\lambda}^s$	-	-	-	-	-0.835	1.066	-1.335	1.446	-1.311	1.439	-0.014	0.080
		$\hat{\lambda}^o$	-	-	-	-	-0.485	0.757	-0.073	0.091	-0.072	0.089	-0.011	0.041
		$\hat{\rho}$	-0.013	0.321	-0.045	0.322	0.080	0.359	-0.190	0.477	-0.188	0.448	-0.008	0.207
		$\hat{\sigma}^2$	2.798	3.103	2.756	3.072	0.843	1.467	0.271	0.449	0.291	0.418	-0.017	0.168

**Table 5:** Percentage biases and RMSEs of marginal effects estimates in the context of the sample selection model with a spatial error in both the selection and outcome equations ( $2n = 300$ ).

$\lambda^s$	$\lambda^o$		MLE		HE		HMLE		GMM		GMM2		PMLE	
			p.bias	RMSE	p.bias	RMSE	p.bias	RMSE	p.bias	RMSE	p.bias	RMSE	p.bias	RMSE
0.00	0.00	$mf x_1$	0.258	0.017	0.206	0.018	0.254	0.017	1.600	0.020	2.400	0.021	0.224	0.017
		$mf x_2$	0.863	0.016	0.934	0.016	0.865	0.016	1.487	0.020	2.543	0.019	0.802	0.016
		$mf x_3$	0.197	0.086	0.241	0.090	0.209	0.086	-0.587	0.104	0.905	0.267	0.184	0.087
		$mf x_4$	0.199	0.050	0.194	0.050	0.195	0.050	0.207	0.051	0.192	0.050	0.176	0.050
0.00	0.40	$mf x_1$	0.723	0.018	0.787	0.018	0.728	0.018	1.911	0.021	2.575	0.023	0.688	0.018
		$mf x_2$	0.784	0.015	0.738	0.015	0.761	0.015	1.377	0.018	2.397	0.021	0.718	0.015
		$mf x_3$	0.485	0.094	0.355	0.097	0.444	0.094	0.743	0.272	1.340	0.260	0.173	0.093
		$mf x_4$	0.000	0.052	0.028	0.052	-0.003	0.052	0.024	0.052	0.082	0.053	-0.010	0.051
0.00	0.85	$mf x_1$	0.399	0.017	0.461	0.018	0.433	0.017	2.403	0.026	3.276	0.025	0.492	0.018
		$mf x_2$	0.449	0.016	0.903	0.016	0.600	0.016	-0.849	0.042	1.091	0.036	0.728	0.016
		$mf x_3$	0.552	0.185	0.541	0.187	0.958	0.173	7.645	0.216	8.684	0.184	0.339	0.132
		$mf x_4$	-0.213	0.100	-0.183	0.100	0.055	0.089	0.327	0.116	0.253	0.102	0.428	0.067
0.40	0.00	$mf x_1$	0.729	0.019	0.820	0.019	0.734	0.019	1.811	0.023	2.744	0.023	0.740	0.019
		$mf x_2$	0.662	0.016	0.572	0.017	0.679	0.016	0.972	0.021	2.057	0.020	0.603	0.016
		$mf x_3$	0.640	0.087	0.352	0.091	0.588	0.087	-0.534	0.102	-0.192	0.097	0.381	0.087
		$mf x_4$	-0.285	0.051	-0.233	0.051	-0.285	0.051	-0.231	0.051	-0.233	0.052	-0.293	0.051
0.40	0.40	$mf x_1$	1.439	0.020	1.531	0.020	1.462	0.020	2.953	0.026	3.629	0.027	1.500	0.019
		$mf x_2$	0.846	0.017	0.738	0.017	0.830	0.017	1.884	0.025	2.603	0.024	0.766	0.017
		$mf x_3$	-0.053	0.092	-0.392	0.097	-0.053	0.092	-1.658	0.198	-0.865	0.192	-0.307	0.091
		$mf x_4$	0.055	0.053	0.083	0.053	0.058	0.053	0.065	0.053	0.081	0.054	0.044	0.052
0.40	0.85	$mf x_1$	0.631	0.019	0.868	0.019	0.756	0.019	2.989	0.032	3.999	0.031	0.809	0.019
		$mf x_2$	0.622	0.018	1.014	0.017	0.676	0.017	-2.015	0.048	0.077	0.044	0.947	0.017
		$mf x_3$	-1.178	0.175	-1.377	0.181	0.437	0.162	8.167	0.193	10.488	0.180	0.436	0.137
		$mf x_4$	-0.240	0.101	-0.163	0.102	-0.359	0.091	0.360	0.122	0.158	0.099	-0.092	0.064
0.85	0.00	$mf x_1$	3.362	0.026	3.324	0.027	3.510	0.026	3.426	0.033	4.961	0.035	3.770	0.024
		$mf x_2$	4.351	0.023	4.470	0.023	4.298	0.023	3.438	0.032	4.893	0.034	4.064	0.021
		$mf x_3$	0.064	0.079	-0.029	0.079	-0.084	0.079	-0.235	0.087	-0.196	0.081	-0.477	0.079
		$mf x_4$	-0.281	0.050	-0.256	0.050	-0.280	0.050	-0.210	0.051	-0.293	0.051	-0.254	0.050
0.85	0.40	$mf x_1$	2.378	0.026	2.323	0.026	2.584	0.026	4.441	0.037	5.290	0.044	3.211	0.024
		$mf x_2$	3.481	0.022	3.371	0.022	3.331	0.022	2.910	0.033	3.550	0.043	2.847	0.021
		$mf x_3$	-0.364	0.082	-0.510	0.083	-0.537	0.083	-0.745	0.087	-0.527	0.087	-0.908	0.081
		$mf x_4$	-0.183	0.054	-0.178	0.054	-0.210	0.054	-0.112	0.054	-0.178	0.056	-0.164	0.053
0.85	0.85	$mf x_1$	2.418	0.026	3.198	0.026	2.876	0.026	7.247	0.046	8.392	0.050	3.637	0.024
		$mf x_2$	2.745	0.024	4.455	0.023	3.145	0.023	-0.716	0.055	2.428	0.053	3.578	0.021
		$mf x_3$	0.154	0.148	-0.497	0.152	1.347	0.143	2.839	0.186	6.733	0.155	0.460	0.115
		$mf x_4$	0.238	0.093	0.345	0.094	0.113	0.084	0.957	0.137	0.848	0.127	0.002	0.061

**Table 6:** AMADs and standard errors of parameter estimates in the context of the sample selection model with a spatial lag in both the selection and outcome equations ( $2n = 300$ ).

$\lambda^s$	$\lambda^o$		$\hat{\beta}_1^s$	$\hat{\beta}_2^s$	$\hat{\beta}_3^s$	$\hat{\beta}_1^o$	$\hat{\beta}_2^o$	$\hat{\beta}_3^o$	$\hat{\lambda}^s$	$\hat{\lambda}^o$	$\hat{\rho}$	$\hat{\sigma}^2$
0.00	0.00	AMAD	0.187	0.135	0.132	0.124	0.095	0.049	0.116	0.085	0.221	0.120
		SE	0.186	0.139	0.134	0.127	0.098	0.051	0.113	0.085	0.226	0.120
0.00	0.40	AMAD	0.180	0.139	0.129	0.098	0.099	0.053	0.110	0.064	0.225	0.120
		SE	0.187	0.138	0.133	0.101	0.099	0.052	0.112	0.063	0.220	0.126
0.00	0.85	AMAD	0.203	0.135	0.132	0.104	0.132	0.070	0.115	0.030	0.235	0.199
		SE	0.188	0.141	0.133	0.100	0.131	0.070	0.114	0.024	0.240	0.200
0.40	0.00	AMAD	0.167	0.147	0.131	0.120	0.094	0.051	0.084	0.088	0.213	0.116
		SE	0.168	0.145	0.135	0.124	0.094	0.050	0.084	0.086	0.222	0.117
0.40	0.40	AMAD	0.169	0.146	0.139	0.099	0.095	0.053	0.085	0.063	0.210	0.123
		SE	0.169	0.144	0.136	0.102	0.096	0.053	0.084	0.062	0.214	0.124
0.40	0.85	AMAD	0.162	0.145	0.130	0.106	0.128	0.070	0.088	0.034	0.220	0.208
		SE	0.171	0.145	0.136	0.101	0.125	0.070	0.086	0.025	0.222	0.198
0.85	0.00	AMAD	0.236	0.253	0.213	0.113	0.080	0.050	0.043	0.082	0.259	0.106
		SE	0.248	0.242	0.219	0.121	0.083	0.052	0.044	0.089	0.263	0.108
0.85	0.40	AMAD	0.216	0.240	0.199	0.097	0.083	0.052	0.044	0.065	0.230	0.112
		SE	0.231	0.234	0.206	0.102	0.084	0.053	0.041	0.065	0.249	0.114
0.85	0.85	AMAD	0.237	0.233	0.216	0.107	0.117	0.079	0.042	0.025	0.245	0.202
		SE	0.240	0.236	0.210	0.101	0.112	0.069	0.039	0.024	0.261	0.170

## Appendix B Some Additional Notation

If  $A$  is a matrix,  $\text{Diag}(A)$  indicates a square diagonal matrix with the diagonal elements of  $A$  on the main diagonal of  $\text{Diag}(A)$ , while  $\text{diag}(A)$  denotes a vector of the diagonal elements in  $A$ . If  $a$  is a vector, then  $\text{Diag}(a)$  indicates a square diagonal matrix with the elements of vector  $a$  on the main diagonal. If  $\tau_1$  and  $\tau_2$  are scalars, then  $\text{Diag}\{\tau_1, \tau_2\}$  denotes a diagonal matrix with  $\tau_1, \tau_2$  on the main diagonal. For some constant  $k$ ,  $\text{Diag}(\cdot)^k := (\text{Diag}(\cdot))^k$ . For some matrix  $A$ ,  $\text{maxeig}(A)$  and  $\text{mineig}(A)$  denote the maximum and minimum eigenvalue of  $A$ , respectively. If  $R$  and  $R_{g,n}^{11}(\theta)$  are correlation matrices, then for notational convenience  $\rho := R_{12}$  and  $\rho_{g,n}^{11}(\theta) := R_{g,n}^{11}(\theta)$ . We use constants  $C_1, C_2, \dots$ , which can be different in different places.

## Appendix C The Likelihood Function and Marginal Effects

### C.1 The likelihood function

There are four scenarios:  $y_{g1,n}^s = 1$  and  $y_{g2,n}^s = 0$ ,  $y_{g1,n}^s = 0$  and  $y_{g2,n}^s = 1$ ,  $y_{g1,n}^s = y_{g2,n}^s = 1$ , and  $y_{g1,n}^s = y_{g2,n}^s = 0$ . We derive the log-likelihood contribution based on the third scenario, while for the other scenarios it can be done in a similar way. Let  $f(\cdot)$  without any index denote a generic density function. Then the Bayes rule and Assumption 2(ii) imply that  $d_{g,n}^{11}f(y_{g,n}^s = \iota_2, y_{g,n}^o | X_n^s, X_n^o) = d_{g,n}^{11}f(y_{g,n}^s = \iota_2, y_{g,n}^{*o} | X_n^s, X_n^o) = P[y_{g,n}^s = \iota_2 | y_{g,n}^{*o}, X_n^s, X_n^o] \cdot d_{g,n}^{11}f(y_{g,n}^{*o} | X_n^o)$ . By Assumption 2,  $y_{g,n}^{*o} | X_n^o \sim \mathcal{N}(S_{g,n}^o(\lambda^o)X_n^o\beta^o, \Omega_{g,n}^{oo}(\theta))$ , thus  $d_{g,n}^{11}f(y_{g,n}^{*o} | X_n^o) = d_{g,n}^{11}\phi_2(y_{g,n}^{*o} - S_{g,n}^o(\lambda^o)X_n^o\beta^o, \Omega_{g,n}^{oo}(\theta)) = d_{g,n}^{11}\phi_2(y_{g,n}^o - S_{g,n}^o(\lambda^o)X_n^o\beta^o, \Omega_{g,n}^{oo}(\theta))$ . Next,

$$\begin{aligned} P[y_{g,n}^s = \iota_2 | y_{g,n}^{*o}, X_n^s, X_n^o] &= P[y_{g,n}^{*s} > 0 | y_{g,n}^{*o}, X_n^s, X_n^o] \\ &= P[S_{g,n}^s(\lambda^s)X_n^s\beta^s + \varepsilon_{g,n}^s(\lambda^s) > 0 | y_{g,n}^{*o}, X_n^s, X_n^o] \\ &= P[-\varepsilon_{g,n}^s(\lambda^s) < S_{g,n}^s(\lambda^s)X_n^s\beta^s | \varepsilon_{g,n}^o(\lambda^o), X_n^s, X_n^o] \\ &= P[-\varepsilon_{g,n}^s(\lambda^s) < \tilde{S}_{g,n}^s(\lambda^s)X_n^s\beta^s | \varepsilon_{g,n}^o(\lambda^o), X_n^s, X_n^o], \end{aligned}$$

where  $\tilde{S}_{g,n}^s(\lambda)$  is defined in Section 3 with  $\zeta_{g,n} = \iota_2$  in this case. Given the definitions of  $\tilde{\Omega}_{g,n}^{ss}(\theta)$ ,  $\tilde{\Omega}_{g,n}^{so}(\theta)$ , and  $\Omega_{g,n}^{oo}(\theta)$  in Section 3 with  $\zeta_{g,n} = \iota_2$ , note that

$$\begin{pmatrix} -\varepsilon_{g,n}^s(\lambda^s) \\ \varepsilon_{g,n}^o(\lambda^o) \end{pmatrix} \bigg| X_n^s, X_n^o \sim \mathcal{N} \left( 0, \begin{pmatrix} \tilde{\Omega}_{g,n}^{ss}(\theta) & \tilde{\Omega}_{g,n}^{so}(\theta) \\ \tilde{\Omega}_{g,n}^{so'}(\theta) & \Omega_{g,n}^{oo}(\theta) \end{pmatrix} \right).$$

Thus,  $-\varepsilon_{g,n}^s(\lambda^s)|\varepsilon_{g,n}^o(\lambda^o), X_n^s, X_n^o \sim \mathcal{N}(\tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)\varepsilon_{g,n}^o(\lambda^o), \Sigma_{g,n}^{11}(\theta))$ , where  $\Sigma_{g,n}^{11}(\theta) = \tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{\Omega}_{g,n}^{so'}(\theta)$ . Substituting for  $\varepsilon_{g,n}^o(\lambda^o)$  from model (3) and interchanging  $y_{g,n}^{*o}$  and  $y_{g,n}^o$  as before, the likelihood contribution equals  $d_{g,n}^{11}P[y_{g,n}^s = \iota_2|y_{g,n}^{*o}, X_n^s, X_n^o] \cdot f(y_{g,n}^{*o}|X_n^o) = d_{g,n}^{11}\Phi_2(\tilde{S}_{g,n}^s(\lambda^s)X_n^s\beta^s - \mu_{g,n}^{11}(\theta), \Sigma_{g,n}^{11}(\theta))\phi_2(y_{g,n}^o - S_{g,n}^o(\lambda^o)X_n^o\beta^o, \Omega_{g,n}^{oo}(\theta))$ , where  $\mu_{g,n}^{11}(\theta) = \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)(y_{g,n}^o - S_{g,n}^o(\lambda^o)X_n^o\beta^o)$ . The result in (4) follows by noting that  $z_{g,n}(\theta) = y_{g,n}^o - S_{g,n}^o(\lambda^o)X_n^o\beta^o$  and  $v_{g,n}^{11}(\theta) = \text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-1/2}(\tilde{S}_{g,n}^s(\lambda^s)X_n^s\beta^s - \mu_{g,n}^{11}(\theta))$ . The log-likelihood contributions based on the other scenarios can be obtained similarly, see (4).

## C.2 Marginal effects

**Spatial lag model.** Let  $P[y^s = 1|X_n^s] = (P[y_{1,n}^s = 1|X_n^s], \dots, P[y_{2n,n}^s = 1|X_n^s])'$ . Then  $\partial P[y^s = 1|X_n^s]/\partial X_{l,n}^{s'} = \text{Diag}(\chi_n(\theta_0))S_n^s(\lambda_0^s)\beta_{0l}^s$  is a matrix of marginal effects associated with a regressor  $l$ , where

$$\chi_n(\theta_0) = \left( \frac{\phi(b_{1,n}(\theta_0))}{\sqrt{\Omega_{11,n}^{ss}(\theta_0)}}, \dots, \frac{\phi(b_{2n,n}(\theta_0))}{\sqrt{\Omega_{2n,2n,n}^{ss}(\theta_0)}} \right)'$$

with

$$b_n(\theta_0) = \left( \frac{S_{1,n}^s(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{11,n}^{ss}(\theta_0)}}, \dots, \frac{S_{2n,n}^s(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{2n,2n,n}^{ss}(\theta_0)}} \right)'.$$

LeSage and Pace (2009) propose to use three types of marginal effects for a spatial lag model: total  $((\partial P[y^s = 1|X_n^s]/\partial X_{l,n}^{s'})\iota_{2n})$ , direct  $(\text{diag}(\partial P[y^s = 1|X_n^s]/\partial X_{l,n}^{s'}))$ , and indirect that is equal to the difference of the first two marginal effects;  $\iota_{2n}$  denotes here the  $2n$ -dimensional vector of ones. The average total, average direct, and average indirect effects are obtained by calculating the averages of these vectors.

Next, note that

$$\begin{aligned} E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_n^s, X_n^o] &= E[S_{i,n}^o(\lambda_0^o)X_n^o\beta_0^o + \varepsilon_{i,n}^o(\lambda_0^o)|y_{i,n}^s = 1, X_n^s, X_n^o] \\ &= S_{i,n}^o(\lambda_0^o)X_n^o\beta_0^o + E[\varepsilon_{i,n}^o(\lambda_0^o)|\varepsilon_{i,n}^s(\lambda_0^s) > -S_{i,n}^s(\lambda_0^s)X_n^s\beta_0^s, X_n^s] \\ &= S_{i,n}^o(\lambda_0^o)X_n^o\beta_0^o + \frac{\Omega_{ii,n}^{so}(\theta_0)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}\Omega_{ii,n}^{oo}(\theta_0)}\sqrt{\Omega_{ii,n}^{oo}(\theta_0)}\frac{\phi\left(-\frac{S_{i,n}^s(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)}{1 - \Phi\left(-\frac{S_{i,n}^s(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)} \\ &= S_{i,n}^o(\lambda_0^o)X_n^o\beta_0^o + \frac{\Omega_{ii,n}^{so}(\theta_0)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\frac{\phi\left(\frac{S_{i,n}^s(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)}{\Phi\left(\frac{S_{i,n}^s(\lambda_0^s)X_n^s\beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}\right)}, \end{aligned}$$

where the third equality follows by Theorem 24.5 of Greene (2008), which states that if  $y$  and  $z$  have a



bivariate normal distribution with means  $\mu_y$  and  $\mu_z$ , standard deviations  $\sigma_y$  and  $\sigma_z$ , and correlation coefficient  $\rho$ , then  $E[y|z > a] = \mu_y + \rho\sigma_y\phi(\alpha_z)/(1 - \Phi(\alpha_z))$  with  $\alpha_z = (a - \mu_z)/\sigma_z$ . Thus, the marginal effect of  $E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_n^s, X_n^o]$  with respect to  $X_{i,n}^o$  depends on whether the explanatory variable is present in both the selection and outcome equations or only in one. Without loss of generality, let the first  $L_1$  explanatory variables be the same in both the selection and outcome equations and ordered in the same way, while the remaining  $L - L_1$  variables be different. Moreover, denote  $E[y^{*o}|y^s = 1, X_n^s, X_n^o] = (E[y_{1,n}^{*o}|y_{1,n}^s = 1, X_n^s, X_n^o], \dots, E[y_{2n,n}^{*o}|y_{2n,n}^s = 1, X_n^s, X_n^o])'$ .

*Case 1.  $l \leq L_1$ :*

Let  $\pi_n(\theta_0) = (\phi(b_{1,n}(\theta_0))/\Phi(b_{1,n}(\theta_0)), \dots, \phi(b_{2n,n}(\theta_0))/\Phi(b_{2n,n}(\theta_0)))'$  and  $\gamma_n(\theta_0) = (\Omega_{11,n}^{so}(\theta_0)/\Omega_{11,n}^{ss}(\theta_0), \dots, \Omega_{2n2n,n}^{so}(\theta_0)/\Omega_{2n2n,n}^{ss}(\theta_0))'$ . Then

$$\frac{\partial E[y^{*o}|y^s = 1, X_n^s, X_n^o]}{\partial X_{i,n}^{o'}} = S_n^o(\lambda_0^o)\beta_{0l}^o - \text{Diag}(\gamma_n(\theta_0)) (\text{Diag}(b_n(\theta_0))\text{Diag}(\pi_n(\theta_0)) + \text{Diag}(\pi_n(\theta_0))^2) S_n^s(\lambda_0^s)\beta_{0l}^s.$$

*Case 2.  $l > L_1$ :*

Now the exogenous variable is present only in the outcome equation, thus the formula simplifies:

$$\frac{\partial E[y^{*o}|y^s = 1, X_n^s, X_n^o]}{\partial X_{i,n}^{o'}} = S_n^o(\lambda_0^o)\beta_{0l}^o.$$

The total, direct, and indirect marginal effects for both cases are obtained analogously to  $\partial P[y^s = 1|X_n^s]/\partial X_{i,n}^{s'}$ .

**Spatial error model.** In the spatial error case, the indirect marginal effects are equal to zero. It is thus enough to consider the marginal effects with respect to “own” exogenous variables:

$$\frac{\partial P[y_{i,n}^s = 1|X_{i,n}^s]}{\partial X_{i,n}^s} = \phi(b_{i,n}^e(\theta_0)) \frac{\beta_{0l}^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}$$

for  $i = 1, \dots, 2n$ , where  $b_{i,n}^e(\theta_0) = X_{i,n}^s\beta_0^s/\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}$ . As before, the marginal effect of  $E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_{i,n}^s, X_{i,n}^o]$  with respect to  $X_{i,n}^o$  depends on whether the explanatory variable is in both equations or only in one.

*Case 1.  $l \leq L_1$ :*

$$\frac{\partial E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_{i,n}^s, X_{i,n}^o]}{\partial X_{i,n}^o} = \beta_{0l}^o - \gamma_{i,n}(\theta_0) \left( b_{i,n}^e(\theta_0)\pi_{i,n}^e(\theta_0) + \pi_{i,n}^{e^2}(\theta_0) \right) \beta_{0l}^s$$

with  $\pi_{i,n}^e(\theta_0) = \phi(b_{i,n}^e(\theta_0))/\Phi(b_{i,n}^e(\theta_0))$ .

Case 2.  $l > L_1$ :

$$\frac{\partial E[y_{i,n}^{*o}|y_{i,n}^s = 1, X_{i,n}^s, X_{i,n}^o]}{\partial X_{il,n}^o} = \beta_{0l}^o.$$

## Appendix D Some Theorems and Technical Lemmas

The appendix contains important theorems as well as several technical lemmas, which will be used later to prove lemmas and theorems in Appendices E and F. The proofs of Lemmas D.5–D.10 are provided in supplementary Appendix H.

**Theorem D.1** (Follows from Theorem 1 of Jenish and Prucha, 2012). *Under Assumption 3, if*

- (i)  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  is uniformly  $L_1$ -NED on an  $\alpha$ -mixing random field  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$ ,
- (ii)  $Z_{g,n}$  is  $L_p$ -bounded uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , for some  $p > 1$ ,
- (iii) the  $\alpha$ -mixing coefficients of the input process  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  satisfy  $\bar{\alpha}(k, m, s) \leq (k + m)^\tau \hat{\alpha}(s)$ ,  $\tau \geq 0$ , with some  $\hat{\alpha}(s) \rightarrow 0$  as  $s \rightarrow \infty$ , such that  $\sum_{s=1}^\infty s^{2\bar{d}-1} \hat{\alpha}(s) < \infty$ ,

then  $n^{-1} \sum_{g \in \mathcal{G}_n} (Z_{g,n} - E[Z_{g,n}]) \xrightarrow{L_1} 0$ .

**Theorem D.2** (Follows from Proposition 3 of Jenish and Prucha, 2012). *Consider transformations of  $Z_{g,n}$  given by a family of functions  $h_{g,n} : \mathbb{R}^{K_Z} \rightarrow \mathbb{R}^{K_h}$ . Suppose that, for all  $(z, z^\bullet) \in \mathbb{R}^{K_Z} \times \mathbb{R}^{K_Z}$  and all  $g \in \mathcal{G}_n$  and  $n \in \mathbb{N}$ ,*

- (i)  $\|h_{g,n}(z) - h_{g,n}(z^\bullet)\| \leq B_{g,n}(z, z^\bullet) \|z - z^\bullet\|$ , where  $B_{g,n}(z, z^\bullet) : \mathbb{R}^{K_Z} \times \mathbb{R}^{K_Z} \rightarrow \mathbb{R}_+$ ,
- (ii)  $\sup_s \|B_{g,n}^{(s)}\|_2 < \infty$ ,
- (iii)  $\sup_s \left\| B_{g,n}^{(s)} \|Z_{g,n} - Z_{g,n}^{(s)}\| \right\|_r < \infty$  for some  $r > 2$ , where  $B_{g,n}^{(s)} = B_{g,n}(Z_{g,n}, Z_{g,n}^{(s)})$  and  $Z_{g,n}^{(s)} = E[Z_{g,n} | \mathcal{F}_{g,n}(s)]$ ,
- (iv)  $\|h_{g,n}(Z_{g,n})\|_2 < \infty$ ,
- (v)  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  is  $L_2$ -NED of size  $-\lambda$  on  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  with scaling factors  $\{t_{g,n}\}_{g \in \mathcal{G}_n}$ .

Then  $h_{g,n}(Z_{g,n})$  is  $L_2$ -NED of size  $-\lambda(r-2)/(2r-2)$  on  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  with scaling factors

$$t_{g,n}^{(r-2)/(2r-2)} \sup_s \|B_{g,n}^{(s)}\|_2^{(r-2)/(2r-2)} \left\| B_{g,n}^{(s)} \|Z_{g,n} - Z_{g,n}^{(s)}\| \right\|_r^{r/(2r-2)}.$$

**Theorem D.3** (Follows from Corollary 1 of Jenish and Prucha, 2012). *For  $K_Z$ -dimensional random vectors  $Z_{g,n}$ ,  $g \in \mathcal{G}_n$ , let  $S_n = \sum_{g \in \mathcal{G}_n} Z_{g,n}$  and  $\Psi_n = \text{Var} \left( \sum_{g \in \mathcal{G}_n} Z_{g,n} \right)$ . Under Assumption 3, if*

- (i)  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  is a zero mean random field,
- (ii)  $Z_{g,n}$  is uniformly  $L_{2+\delta}$ -bounded, for some  $\delta > 0$ ,
- (iii)  $\{Z_{g,n}\}_{g \in \mathcal{G}_n}$  is  $L_2$ -NED random field on an  $\alpha$ -mixing random field  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  with NED coefficients  $\psi(s)$  and NED scaling factors  $\{t_{g,n}\}_{g \in \mathcal{G}_n}$ ,
- (iv) NED coefficients satisfy  $\sum_{s=1}^{\infty} s^{2\tilde{d}-1} \psi(s) < \infty$ ,
- (v) NED scaling factors satisfy  $\sup_{n,g} t_{g,n} < \infty$ ,
- (vi) the  $\alpha$ -mixing coefficients of  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  satisfy  $\bar{\alpha}(k, m, s) \leq (k+m)^\tau \hat{\alpha}(s)$ , for some  $\tau \geq 0$  and  $\hat{\alpha}(s) \rightarrow 0$  as  $s \rightarrow \infty$ , such that for some  $\delta > 0$ ,  $\sum_{s=1}^{\infty} s^{2\tilde{d}(\tau_*(+1)-1) \hat{\alpha}^{\delta/(4+2\delta)}(s)} < \infty$ , where  $\tau_* = \delta\tau/(2+\delta)$ ,
- (vii)  $\inf_n \frac{1}{n} \text{mineig}(\Psi_n) > 0$ ,

then  $\Psi_n^{-1/2} S_n \xrightarrow{d} \mathcal{N}(0, I_{K_Z})$  as  $n \rightarrow \infty$ .

**Theorem D.4** (Follows from Theorem 17.8 of Davidson, 1994). *Let for some  $p \geq 1$ ,  $\|X_{g,n} - E[X_{g,n} | \mathcal{F}_{g,n}(s)]\|_p \leq t_{g,n}^X \psi^X(s)$  and  $\|Y_{g,n} - E[Y_{g,n} | \mathcal{F}_{g,n}(s)]\|_p \leq t_{g,n}^Y \psi^Y(s)$ . Then  $\|X_{g,n} + Y_{g,n} - E[X_{g,n} + Y_{g,n} | \mathcal{F}_{g,n}(s)]\|_p \leq t_{g,n} \psi(s)$ , where  $t_{g,n} = \max\{t_{g,n}^X, t_{g,n}^Y\}$  and  $\psi(s) = \psi^X(s) + \psi^Y(s)$ .*

*Specifically, if  $\{X_{g,n}\}_{g \in \mathcal{G}_n}$  and  $\{Y_{g,n}\}_{g \in \mathcal{G}_n}$  are uniform  $L_p$ -NED random fields, then  $\{X_{g,n} + Y_{g,n}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_p$ -NED random field as well.*

**Lemma D.5.** *Let  $v := v(\theta)$  be a 2-dimensional vector,  $R := R(\theta)$  be a  $2 \times 2$  dimensional correlation matrix with the off-diagonal element  $\rho := \rho(\theta)$ ,  $|\rho| < 1$ , and  $P(z, R) = z' R^{-1}(\theta) z$ , where  $z$  is a 2-dimensional vector. Then*

$$\begin{aligned}
\frac{\partial \ln \Phi_2(v, R)}{\partial v} &= \xi(v, R), \\
\frac{\partial \ln \Phi_2(v, R)}{\partial \theta} &= \xi_1(v, R) \frac{\partial v_1}{\partial \theta} + \xi_2(v, R) \frac{\partial v_2}{\partial \theta} - \frac{1}{2} \left( \frac{\partial \ln |R|}{\partial \theta} + E_V \left[ \frac{\partial P(V, R)}{\partial \theta} \middle| V \leq v \right] \right), \\
\frac{\partial^2 \ln \Phi_2(v, R)}{\partial \theta \partial \theta'} &= \xi_1(v, R) \frac{\partial^2 v_1}{\partial \theta \partial \theta'} + \xi_2(v, R) \frac{\partial^2 v_2}{\partial \theta \partial \theta'} + \kappa(v, R) \left( \frac{\partial v_1}{\partial \theta} \frac{\partial v_2}{\partial \theta'} + \frac{\partial v_2}{\partial \theta} \frac{\partial v_1}{\partial \theta'} \right) \\
&\quad - \frac{1}{2} (A(v, R) + B(v, R) + A'(v, R) + B'(v, R) + E_V[G(V, R) | V \leq v]) \\
&\quad - \frac{1}{\Phi_2^2(v, R)} \frac{\partial \Phi_2(v, R)}{\partial \theta} \frac{\partial \Phi_2(v, R)}{\partial \theta'},
\end{aligned}$$

where (keeping the dependence on  $\theta$  implicit)

$$\xi(v, R) = (\xi_1(v, R), \xi_2(v, R))' = \left( \frac{\phi(v_1)\Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)}, \frac{\phi(v_2)\Phi\left(\frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} \right)', \quad (\text{D.1})$$

$$\kappa(v, R) = \frac{\phi_2(v, R)}{\Phi_2(v, R)}, \quad (\text{D.2})$$

$$\begin{aligned} A(v, R) &= \xi_1(v, R) \frac{\partial v_1}{\partial \theta} \left( \frac{\partial \ln |R|}{\partial \theta'} + E_{\tilde{V}_2} \left[ \frac{\partial P((v_1, \tilde{V}_2)', R)}{\partial \theta'} \middle| \tilde{V}_2 \leq v_2 \right] \right), \\ B(v, R) &= \xi_2(v, R) \frac{\partial v_2}{\partial \theta} \left( \frac{\partial \ln |R|}{\partial \theta'} + E_{\tilde{V}_1} \left[ \frac{\partial P((\tilde{V}_1, v_2)', R)}{\partial \theta'} \middle| \tilde{V}_1 \leq v_1 \right] \right), \\ G(z, R) &= \frac{\partial^2 \ln |R|}{\partial \theta \partial \theta'} + \frac{\partial^2 P(z, R)}{\partial \theta \partial \theta'} - \frac{1}{2} \left( \frac{\partial \ln |R|}{\partial \theta} + \frac{\partial P(z, R)}{\partial \theta} \right) \left( \frac{\partial \ln |R|}{\partial \theta'} + \frac{\partial P(z, R)}{\partial \theta'} \right) \text{ and} \\ V &\sim \mathcal{N}(0, R), \quad \tilde{V}_1 \sim \mathcal{N}(\rho v_2, 1 - \rho^2), \quad \tilde{V}_2 \sim \mathcal{N}(\rho v_1, 1 - \rho^2). \end{aligned} \quad (\text{D.3})$$

**Lemma D.6.** *Let  $v$  and  $R$  be a 2-dimensional vector and a  $2 \times 2$  dimensional correlation matrix with the off-diagonal element  $\rho$ ,  $|\rho| < 1$ , respectively. Then for some constants  $C_1, C_2, C_3 > 0$ ,*

$$\begin{aligned} \frac{\phi(v_1)\Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} &\leq C_1(1 - |\rho|)^{-7} \left( (|v_1| + |v_2| + C_2)^8 + \left(1 - \Phi\left((1 - |\rho|)^{-1/2}\right)\right)^{-2} \right), \\ \frac{\phi_2(v, R)}{\Phi_2(v, R)} &\leq C_1(1 - |\rho|)^{-3} \left( (|v_1| + |v_2| + C_2)^2 + \left(1 - \Phi\left((1 - |\rho|)^{-1/2}\right)\right)^{-2} \right), \\ \Phi_2(0, R) &\geq C_3(1 - |\rho|)^{1/2}. \end{aligned}$$

**Lemma D.7.** *Let  $\theta$  be a  $p \times 1$  vector,  $f(\theta)$  an  $n \times 1$  vector, and  $F(\theta)$  an  $n \times n$  symmetric matrix. Then*

$$\begin{aligned} \frac{\partial |F(\theta)|}{\partial \theta} &= |F(\theta)| K(\theta) \text{vec } F^{-1}(\theta), \\ \frac{\partial^2 |F(\theta)|}{\partial \theta \partial \theta'} &= K(\theta) \text{vec } F^{-1}(\theta) \frac{\partial |F(\theta)|}{\partial \theta'} + |F(\theta)| ((\text{vec } F^{-1}(\theta))' \otimes I_p) \frac{\partial \text{vec } K(\theta)}{\partial \theta'} \\ &\quad + |F(\theta)| K(\theta) \frac{\partial \text{vec } F^{-1}(\theta)}{\partial \theta'}, \\ \frac{\partial (f'(\theta) F^{-1}(\theta) f(\theta))}{\partial \theta} &= 2L(\theta) F^{-1}(\theta) f(\theta) + M(\theta) (f(\theta) \otimes f(\theta)), \\ \frac{\partial^2 (f'(\theta) F^{-1}(\theta) f(\theta))}{\partial \theta \partial \theta'} &= 2(f'(\theta) F^{-1}(\theta) \otimes I_p) \frac{\partial \text{vec } L(\theta)}{\partial \theta'} + 2(f'(\theta) \otimes L(\theta)) \frac{\partial \text{vec } F^{-1}(\theta)}{\partial \theta'} \\ &\quad + (f'(\theta) \otimes f'(\theta) \otimes I_p) \frac{\partial \text{vec } M(\theta)}{\partial \theta'} \\ &\quad + (2L(\theta) F^{-1}(\theta) + M(\theta) (K_{1n} \otimes I_n) [(I_n \otimes f(\theta)) + (f(\theta) \otimes I_n)]) \frac{\partial f(\theta)}{\partial \theta'}, \end{aligned}$$

where

$$K(\theta) = \left( \frac{\partial \text{vec } F(\theta)}{\partial \theta'} \right)', \quad L(\theta) = \left( \frac{\partial f(\theta)}{\partial \theta'} \right)', \quad M(\theta) = \left( \frac{\partial \text{vec } F^{-1}(\theta)}{\partial \theta'} \right)',$$

and  $K_{1n}$  is the commutation matrix.<sup>16</sup>

**Lemma D.8.** *Let  $A$  and  $B$  be  $m \times n$  and  $p \times q$  matrices. Then  $\|A \otimes B\| = \|A\| \|B\|$ .*

**Lemma D.9.** *Let  $X \sim \mathcal{N}(0, R)$ , where  $R$  is a  $2 \times 2$  dimensional correlation matrix with the off-diagonal element  $\rho$ ,  $|\rho| < 1$ . Then for a 2-dimensional vector of constants  $v = (v_1, v_2)'$ ,*

$$E[XX'|X \leq v] = R - v_1 \xi_1(v, R) A_1(R) - v_2 \xi_2(v, R) A_2(R) + (1 - \rho^2) \kappa(v, R) A_3(R),$$

where  $\xi(v, R)$  and  $\kappa(v, R)$  are defined in (D.1) and (D.2), respectively, and

$$A_1(R) = \begin{pmatrix} 1 & \rho \\ \rho & \rho^2 \end{pmatrix}, \quad A_2(R) = \begin{pmatrix} \rho^2 & \rho \\ \rho & 1 \end{pmatrix}, \quad A_3(R) = \begin{pmatrix} \rho & 1 \\ 1 & \rho \end{pmatrix}. \quad (\text{D.4})$$

**Lemma D.10.** *If for some  $p \geq 1$ ,  $\|X_{i,n} - E[X_{i,n}|\mathcal{F}_{i,n}(s)]\|_{2p} \leq t_{i,n}^X \psi^X(s)$  and  $\|Y_{i,n} - E[Y_{i,n}|\mathcal{F}_{i,n}(s)]\|_{2p} \leq t_{i,n}^Y \psi^Y(s)$ , then  $\|X_{i,n}Y_{i,n} - E[X_{i,n}Y_{i,n}|\mathcal{F}_{i,n}(s)]\|_p \leq t_{i,n} \psi(s)$ , where  $t_{i,n} = \max\{\|X_{i,n}\|_{2p} t_{i,n}^Y, \|Y_{i,n}\|_{2p} t_{i,n}^X, t_{i,n}^X t_{i,n}^Y\}$  and  $\psi(s) = \psi^X(s) + \psi^Y(s) + \psi^X(s)\psi^Y(s)$ . Specifically, if  $\{X_{i,n}\}_{i=1}^n$  and  $\{Y_{i,n}\}_{i=1}^n$  are uniformly  $L_{2p}$ -NED, then  $\{X_{i,n}Y_{i,n}\}_{i=1}^n$  is uniformly  $L_p$ -NED.*

## Appendix E Some Useful Lemmas

The appendix contains several lemmas that establish the uniform ( $L_p$ -) bounds and the NED property of the random variables in the studied sample selection models. The proofs of Lemmas E.1–E.5 are provided in supplementary Appendix I.

**Lemma E.1.**

- (i) *Under Assumptions 1(ii), 2(i), 6, and 7,  $\inf_{n,g} \inf_{\theta \in \Theta} |\Omega_{g,n}^b(\theta)| > 0$  and  $\inf_{n,i} \inf_{\theta \in \Theta} \Omega_{ii,n}^b(\theta) > 0$ , where  $b \in \{ss, oo\}$ .*
- (ii) *Under Assumptions 1(ii) and 7,  $\|\Omega_{g,n}^c(\theta)\|$  and  $\|\partial \text{vec } \Omega_{g,n}^c(\theta)/\partial \theta'\|$  are uniformly bounded in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$ , where  $c \in \{ss, so, oo\}$ .*

---

<sup>16</sup>Let  $A$  be an  $m \times n$  matrix. Then there exists a unique  $mn \times mn$  permutation matrix which transforms  $\text{vec } A$  into  $\text{vec } A'$ , i.e.  $K_{mn} \text{vec } A = \text{vec } A'$ .

(iii) Under Assumptions 1(ii), 6, and 7,  $\|\Omega_{g,n}^{b-1}(\theta)\|$ ,  $\|\partial \text{vec } \Omega_{g,n}^{b-1}(\theta)/\partial \theta'\|$ , and  $\|\partial |\Omega_{g,n}^b(\theta)|/\partial \theta\|$  are uniformly bounded in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$ , where  $b \in \{ss, oo\}$ .

**Lemma E.2.**

- (i) Under Assumption 6,  $\inf_{n,g} \inf_{\theta \in \Theta} \min_{j=1,2} \Sigma_{gj,n}^{11}(\theta) > 0$  and  $\sup_{n,g} \sup_{\theta \in \Theta} |\rho_{g,n}^{11}(\theta)| < 1$ .
- (ii) Under Assumptions 1(ii), 6, and 7,  $\|R_{g,n}^{11}(\theta)\|$ ,  $\|\partial |R_{g,n}^{11}(\theta)|/\partial \theta'\|$ , and  $\|\partial \text{vec } R_{g,n}^{11-1}(\theta)/\partial \theta'\|$  are uniformly bounded in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$ .

**Lemma E.3.** Under Assumptions 1(ii), 2(i), 4(ii), 6, and 7,  $\sup_{\theta \in \Theta} \|S_{g,n}^b(\lambda^b) X_n^b \beta^b\|$ ,  $b \in \{s, o\}$ ,  $\sup_{\theta \in \Theta} \|z_{g,n}(\theta)\|$ ,  $\sup_{\theta \in \Theta} \|\mu_{g,n}^{11}(\theta)\|$ , and  $\sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\|$  are  $L_p$ -bounded uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , for any given  $p \geq 1$ .

**Lemma E.4.** Under Assumptions 1(ii), 2(i), 4(ii), 6, and 7,  $\sup_{\theta \in \Theta} \|\partial z_{g,n}(\theta)/\partial \theta'\|$  and  $\sup_{\theta \in \Theta} \|\partial v_{g,n}^{11}(\theta)/\partial \theta'\|$  are  $L_p$ -bounded uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , for any given  $p \geq 1$ .

**Lemma E.5.** Under Assumptions 1(ii), 2(i), 4(ii), 5, and 7,  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$ ,  $\{z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$ , and  $\{v_{g,n}^{11}(\theta)\}_{g \in \mathcal{G}_n}$  are uniformly  $L_2$ -NED on random field  $\{\eta_{g,n} = (X_{g,n}^s, X_{g,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g \in \mathcal{G}_n}$  with NED coefficients  $\psi^{1/6}(s)$ , where  $\psi(s)$  is defined in Assumption 5.

## Appendix F Proofs of the Asymptotic Results

**Proof of Theorem 1:** We apply Theorem 2.1 of Newey and McFadden (1994). Let  $Q_0(\theta) = \lim_{n \rightarrow \infty} E[Q_n(\theta)]$ .

It is sufficient to verify that (i)  $Q_0(\theta)$  is uniquely maximized at  $\theta_0$ , (ii)  $\Theta$  is compact, (iii)  $Q_0(\theta)$  is continuous, and (iv)  $Q_n(\theta)$  converges uniformly in probability to  $Q_0(\theta)$ . We have already assumed the first two conditions (Assumptions 7 and 8), thus it remains to show that the last two conditions are satisfied. In order to prove uniform convergence in (iv), we apply Theorem 2 of Jenish and Prucha (2009), which requires the uniform  $L_p$ -boundedness (LB),  $p > 1$ , and  $L_0$ -stochastic equicontinuity (SE) of the individual likelihood terms as well as the pointwise convergence (PC) in probability; see the following paragraphs. As the bounds constructed to verify the uniform  $L_p$ -boundedness and  $L_0$ -stochastic equicontinuity are uniform in  $g \in \mathcal{G}_n$ ,  $n \in \mathbb{N}$ , and  $\theta \in \Theta$ , it follows that the whole likelihood function  $Q_n(\theta)$  also satisfies these conditions (once they are verified). The  $L_0$ -stochastic equicontinuity along with the uniform convergence verified for point (iv) below will then imply condition (iii), that is, the continuity of  $Q_0(\theta)$ , and therefore, will allow us to apply Theorem 2.1 of Newey and McFadden (1994) and to claim the consistency of PMLE.

LB: proof of  $\sup_{n,g} E \left[ \sup_{\theta \in \Theta} \left| \sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta) \right| \right]^p < \infty$ , for some  $p > 1$

---

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} \left| \sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta) \right| \right]^p &\leq E \left[ \sum_{a \in \mathcal{A}} \sup_{\theta \in \Theta} |d_{g,n}^a f_{g,n}^a(\theta)| \right]^p \leq 4^{p-1} \sum_{a \in \mathcal{A}} E \left[ \sup_{\theta \in \Theta} |d_{g,n}^a f_{g,n}^a(\theta)| \right]^p \\ &\leq 4^{p-1} \sum_{a \in \mathcal{A}} E \left[ \sup_{\theta \in \Theta} |f_{g,n}^a(\theta)| \right]^p \end{aligned} \quad (\text{F.1})$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , where the first and second inequalities follow by the triangle and Loève's  $c_r$ -inequalities, respectively, whereas the last inequality follows by noting that  $d_{g,n}^a \in \{0, 1\}$ .

We will show that  $\sup_{n,g} E[\sup_{\theta \in \Theta} |f_{g,n}^{11}(\theta)|]^p < \infty$ , while the boundedness of the other terms can be proven in a similar way. By the definitions of  $f_{g,n}^{11}(\theta)$  in (4) and of the multivariate normal density function,

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} |f_{g,n}^{11}(\theta)| \right]^p &\leq 4^{p-1} \left( |\ln 2\pi|^p + \sup_{\theta \in \Theta} |\ln |\Omega_{g,n}^{oo}(\theta)||^p + \sup_{\theta \in \Theta} \|\Omega_{g,n}^{oo-1}(\theta)\|^p E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^{2p} \right. \\ &\quad \left. + E \left[ \sup_{\theta \in \Theta} |\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))| \right]^p \right), \end{aligned} \quad (\text{F.2})$$

where the result follows by the triangle and Loève's  $c_r$ -inequalities. The second and third terms are uniformly bounded by Lemmas E.1 and E.3. Hence, only the last term has to be shown to be uniformly bounded. Let  $\xi(\cdot)$  be defined in the same way as in (D.1), Lemma D.5, with correlation matrix  $R_{g,n}^{11}(\theta)$  and correlation coefficient  $\rho_{g,n}^{11}(\theta)$ , which is the off-diagonal element of  $R_{g,n}^{11}(\theta)$ . Then by the elementwise mean value theorem, there exists  $\tilde{v}_{g,n}^{11}(\theta)$  with elements between 0 and  $v_{g,n}^{11}(\theta)$  and constants  $C_1, \dots, C_8 > 0$  with  $C_3, C_6 \geq 1$  such that

$$\begin{aligned} &|\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))| \\ &\leq |\ln \Phi_2(0, R_{g,n}^{11}(\theta))| + \left| \frac{\partial \ln \Phi_2(\tilde{v}_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial v'} v_{g,n}^{11}(\theta) \right| \\ &= |\ln \Phi_2(0, R_{g,n}^{11}(\theta))| + |\xi'(\tilde{v}_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) v_{g,n}^{11}(\theta)| \\ &\leq |\ln \Phi_2(0, R_{g,n}^{11}(\theta))| + \sum_{j=1}^2 \xi_j(\tilde{v}_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) |v_{gj,n}^{11}(\theta)| \\ &\leq |\ln(C_1(1 - |\rho_{g,n}^{11}(\theta)|)^{1/2})| \\ &\quad + C_2(1 - |\rho_{g,n}^{11}(\theta)|)^{-7} \left( (|\tilde{v}_{g1,n}^{11}(\theta)| + |\tilde{v}_{g2,n}^{11}(\theta)| + C_3)^8 + \left( 1 - \Phi \left( (1 - |\rho_{g,n}^{11}(\theta)|)^{-1/2} \right) \right)^{-2} \right) \sum_{j=1}^2 |v_{gj,n}^{11}(\theta)| \\ &\leq C_4 + C_5 \left( (|\tilde{v}_{g1,n}^{11}(\theta)| + |\tilde{v}_{g2,n}^{11}(\theta)| + C_3)^8 + C_6 \right) \sum_{j=1}^2 |v_{gj,n}^{11}(\theta)| \end{aligned} \quad (\text{F.3})$$

$$\begin{aligned}
&\leq C_4 + C_5 \left( (|v_{g1,n}^{11}(\theta)| + |v_{g2,n}^{11}(\theta)| + C_3)^8 + C_6 \right) \sum_{j=1}^2 |v_{gj,n}^{11}(\theta)| \\
&\leq C_4 + 2C_5 C_6 (|v_{g1,n}^{11}(\theta)| + |v_{g2,n}^{11}(\theta)| + C_3)^9 \\
&= C_4 + 2C_5 C_6 (\|v_{g,n}^{11}(\theta)\|_1 + C_3)^9 \\
&\leq C_4 + 2^9 C_5 C_6 (\|v_{g,n}^{11}(\theta)\|_1^9 + C_3^9) \leq C_7 + C_8 \|v_{g,n}^{11}(\theta)\|^9,
\end{aligned}$$

where Lemma D.5 implies the first equality, the third inequality is implied by Lemma D.6, whereas the fourth inequality follows from Lemma E.2. The conclusion follows by the equivalence of vector norms on finite dimensional vector spaces. Given this result,

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} |\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))| \right]^p &\leq E \left[ \sup_{\theta \in \Theta} (C_7 + C_8 \|v_{g,n}^{11}(\theta)\|^9) \right]^p \\
&\leq 2^{p-1} \left( C_7^p + C_8^p E \left[ \sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\| \right]^{9p} \right) < \infty
\end{aligned} \tag{F.4}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  by Lemma E.3.

SE: proof that  $\sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta)$  is  $L_0$ -stochastically equicontinuous

The  $L_0$ -stochastic equicontinuity will be verified using Proposition 1 of Jenish and Prucha (2009). To apply it, we have to show that the individual likelihood terms are Lipschitz functions in parameters: for any  $\theta, \theta^\bullet \in \Theta$ ,

$$\begin{aligned}
\left| \sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta) - \sum_{a \in \mathcal{A}} d_{g,n}^a f_{g,n}^a(\theta^\bullet) \right| &\leq \sum_{a \in \mathcal{A}} d_{g,n}^a |f_{g,n}^a(\theta) - f_{g,n}^a(\theta^\bullet)| \leq \sum_{a \in \mathcal{A}} |f_{g,n}^a(\theta) - f_{g,n}^a(\theta^\bullet)| \\
&\leq \sum_{a \in \mathcal{A}} \left\| \frac{\partial f_{g,n}^a(\tilde{\theta})}{\partial \theta} \right\| \|\theta - \theta^\bullet\| \leq \sum_{a \in \mathcal{A}} \sup_{\theta \in \Theta} \left\| \frac{\partial f_{g,n}^a(\theta)}{\partial \theta} \right\| \cdot \|\theta - \theta^\bullet\|,
\end{aligned}$$

where we used the elementwise mean value theorem with elements of  $\tilde{\theta}$  being between elements of  $\theta$  and  $\theta^\bullet$ . It thus suffices to show that  $\sup_{n,g} E \left[ \sum_{a \in \mathcal{A}} \sup_{\theta \in \Theta} \|\partial f_{g,n}^a(\theta)/\partial \theta\| \right]^p < \infty$  for some  $p \geq 1$ . Similarly to (F.1), Loève's  $c_r$ -inequality implies that it is enough to prove that all individual terms are bounded, that is,  $\sup_{n,g} E \left[ \sup_{\theta \in \Theta} \|\partial f_{g,n}^a(\theta)/\partial \theta\| \right]^p < \infty$  for all  $a \in \mathcal{A}$ . As before, we establish this result for  $\partial f_{g,n}^{11}(\theta)/\partial \theta$ , while the boundedness of the other terms can be proven in a similar way:

$$\begin{aligned}
&E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial f_{g,n}^{11}(\theta)}{\partial \theta} \right\| \right]^p \\
&= E \left[ \sup_{\theta \in \Theta} \left\| -\frac{1}{2} \left( \frac{1}{|\Omega_{g,n}^{oo}(\theta)|} \frac{\partial |\Omega_{g,n}^{oo}(\theta)|}{\partial \theta} + \frac{\partial (z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta))}{\partial \theta} \right) + \frac{\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta} \right\| \right]^p
\end{aligned}$$



$$\begin{aligned}
&\leq 3^{p-1} \left( \sup_{\theta \in \Theta} \left( \frac{1}{|\Omega_{g,n}^{oo}(\theta)|} \left\| \frac{\partial |\Omega_{g,n}^{oo}(\theta)|}{\partial \theta} \right\| \right)^p + E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial(z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta))}{\partial \theta} \right\| \right]^p \right. \\
&\quad \left. + E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta} \right\| \right]^p \right). \tag{F.5}
\end{aligned}$$

The first term on the right hand side is uniformly bounded by Lemma E.1. To bound the second term on the right hand side of (F.5), we apply Lemma D.7:

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial(z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta))}{\partial \theta} \right\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left\| 2 \left( \frac{\partial z_{g,n}(\theta)}{\partial \theta'} \right)' \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta) \right. \right. \\
&\quad \left. \left. + \left( \frac{\partial \text{vec} \Omega_{g,n}^{oo-1}(\theta)}{\partial \theta'} \right)' (z_{g,n}(\theta) \otimes z_{g,n}(\theta)) \right\| \right]^p \\
&\leq 2^{2p-1} \left( \sup_{\theta \in \Theta} \|\Omega_{g,n}^{oo-1}(\theta)\|^p E \left[ \sup_{\theta \in \Theta} \left( \left\| \frac{\partial z_{g,n}(\theta)}{\partial \theta'} \right\| \|z_{g,n}(\theta)\| \right) \right]^p \right. \\
&\quad \left. + \sup_{\theta \in \Theta} \left\| \frac{\partial \text{vec} \Omega_{g,n}^{oo-1}(\theta)}{\partial \theta'} \right\|^p E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^{2p} \right) < \infty \tag{F.6}
\end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , where the first inequality is implied by Lemma D.8 and Loève's  $c_r$ -inequality. The conclusion follows by applying the Cauchy-Schwartz inequality to  $E [\sup_{\theta \in \Theta} (\|\partial z_{g,n}(\theta)/\partial \theta'\| \|z_{g,n}(\theta)\|)]^p$  and observing that  $E [\sup_{\theta \in \Theta} \|\partial z_{g,n}(\theta)/\partial \theta'\|]^{2p}$  and  $E [\sup_{\theta \in \Theta} \|z_{g,n}(\theta)\|]^{2p}$  are uniformly bounded by Lemmas E.4 and E.3, respectively, while the norms of  $\Omega_{g,n}^{oo-1}(\theta)$  and  $\partial \text{vec} \Omega_{g,n}^{oo-1}(\theta)/\partial \theta'$  are uniformly bounded by Lemma E.1.

Finally, by Lemma D.5, the last term in (F.5) can be bounded (all symbols defined in Lemma D.5 are again indexed by the subscripts  $g$  and  $n$  and superscript 11):

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta} \right\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left\| \sum_{j=1}^2 \xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \frac{\partial v_{gj,n}^{11}(\theta)}{\partial \theta} \right. \right. \\
&\quad \left. \left. - \frac{1}{2} \left( \frac{\partial \ln |R_{g,n}^{11}(\theta)|}{\partial \theta} + E_V \left[ \frac{\partial P(V, R_{g,n}^{11}(\theta))}{\partial \theta} \middle| V \leq v_{g,n}^{11}(\theta) \right] \right) \right\| \right]^p \\
&\leq 4^{p-1} \left( \sum_{j=1}^2 E \left[ \sup_{\theta \in \Theta} \left\| \xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \frac{\partial v_{gj,n}^{11}(\theta)}{\partial \theta} \right\| \right]^p \right. \\
&\quad \left. + \sup_{\theta \in \Theta} \left\| \frac{1}{|R_{g,n}^{11}(\theta)|} \frac{\partial |R_{g,n}^{11}(\theta)|}{\partial \theta} \right\|^p + E \left[ \sup_{\theta \in \Theta} \left\| E_V \left[ \frac{\partial P(V, R_{g,n}^{11}(\theta))}{\partial \theta} \middle| V \leq v_{g,n}^{11}(\theta) \right] \right\| \right]^p \right) \tag{F.7}
\end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , where  $P(v, R_{g,n}^{11}(\theta)) = v' R_{g,n}^{11-1}(\theta) v$  and  $V \sim \mathcal{N}(0, R_{g,n}^{11}(\theta))$ .<sup>17</sup>

<sup>17</sup>We use  $V$  instead of  $V_{g,n}^{11}(\theta)$  in order to simplify the notation.

Now we will prove that each term in (F.7) is uniformly bounded. By the Cauchy-Schwartz inequality,

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} \left\| \xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \frac{\partial v_{gj,n}^{11}(\theta)}{\partial \theta} \right\| \right]^p &\leq \sqrt{E \left[ \sup_{\theta \in \Theta} \xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \right]^{2p}} E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial v_{gj,n}^{11}(\theta)}{\partial \theta'} \right\| \right]^{2p} \\ &\leq \sqrt{E \left[ \sup_{\theta \in \Theta} \xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \right]^{2p}} E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial v_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right]^{2p}, \end{aligned} \quad (\text{F.8})$$

for  $j = 1, 2$ . In the same way as in (F.3),  $\xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \leq C_1 + C_2 \|v_{g,n}^{11}(\theta)\|^8$  for some constants  $C_1, C_2 > 0$ .<sup>18</sup> Thus for (F.8) to be uniformly bounded, it is enough to show that  $E[\sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\|]^{16p}$  and  $E[\sup_{\theta \in \Theta} \|\partial v_{g,n}^{11}(\theta)/\partial \theta'\|]^{2p}$  are uniformly bounded, which is the case by Lemmas E.3 and E.4, respectively. The second term on the right hand side of (F.7) is uniformly bounded because  $\inf_{n,g} |R_{g,n}^{11}(\theta)| = \inf_{n,g} (1 - \rho_{g,n}^{11^2}(\theta)) > 0$  and  $\sup_{n,g} \sup_{\theta \in \Theta} \|\partial R_{g,n}^{11}(\theta)/\partial \theta\| < \infty$  by Lemma E.2. Regarding the last term in (F.7), it is not difficult to see from Lemma D.7 on the first order derivative of a quadratic form and from Lemma E.2 uniformly bounding  $\partial \text{vec } R_{g,n}^{11^{-1}}(\theta)/\partial \theta'$  that

$$\begin{aligned} &\left\| E_V \left[ \frac{\partial P(V, R_{g,n}^{11}(\theta))}{\partial \theta} | V \leq v_{g,n}^{11}(\theta) \right] \right\| \\ &= \left\| E_V \left[ \left( \frac{\partial \text{vec } R_{g,n}^{11^{-1}}(\theta)}{\partial \theta'} \right)' (V \otimes V) | V \leq v_{g,n}^{11}(\theta) \right] \right\| \\ &\leq \left\| \frac{\partial \text{vec } R_{g,n}^{11^{-1}}(\theta)}{\partial \theta'} \right\| \|E_V[V \otimes V | V \leq v_{g,n}^{11}(\theta)]\| \\ &= \left\| \frac{\partial \text{vec } R_{g,n}^{11^{-1}}(\theta)}{\partial \theta'} \right\| \|E_V[VV' | V \leq v_{g,n}^{11}(\theta)]\| \\ &\leq C_3 \|E_V[VV' | V \leq v_{g,n}^{11}(\theta)]\| \\ &= C_3 \|R_{g,n}^{11}(\theta) - v_{g1,n}^{11}(\theta)\xi_1(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))A_1(R_{g,n}^{11}(\theta)) - v_{g2,n}^{11}(\theta)\xi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))A_2(R_{g,n}^{11}(\theta)) \\ &\quad + (1 - \rho_{g,n}^{11^2}(\theta))\kappa(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))A_3(R_{g,n}^{11}(\theta))\| \\ &\leq C_4 [C_5 + |v_{g1,n}^{11}(\theta)\xi_1(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))| + |v_{g2,n}^{11}(\theta)\xi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))| + \kappa(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))] \end{aligned} \quad (\text{F.9})$$

for some constants  $C_3, C_4, C_5 > 0$ , where the last equality follows by Lemma D.9 and its notation: recall that  $V \sim \mathcal{N}(0, R_{g,n}^{11}(\theta))$  and  $2 \times 2$  matrices  $A_1(R_{g,n}^{11}(\theta))$ ,  $A_2(R_{g,n}^{11}(\theta))$ , and  $A_3(R_{g,n}^{11}(\theta))$  are functions of  $\rho_{g,n}^{11}(\theta)$  defined in the same way as in (D.4) and  $\kappa(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) = \phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))/\Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))$ . It remains to show that the supremum of the last expression in (F.9) with respect to  $\theta \in \Theta$  is uniformly  $L_p$ -bounded. By Loève's  $c_r$ -inequality, it suffices to show that  $\sup_{n,g} E [\sup_{\theta \in \Theta} |v_{gj,n}^{11}(\theta)\xi_j(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))|]^p <$

<sup>18</sup>Note that the numbering of constants is renewed for each part of the proof.

$\infty$ ,  $j = 1, 2$ , which is however implied by results in (F.3) and (F.4), and  $\sup_{n,g} E [\sup_{\theta \in \Theta} \kappa(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))]^p < \infty$ . For this last term, Lemmas D.6 and E.2 imply there are constants  $C_6, \dots, C_9 > 0$  such that

$$\begin{aligned} & E \left[ \sup_{\theta \in \Theta} \kappa(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \right]^p \\ & \leq E \left[ \sup_{\theta \in \Theta} \left( C_6 (1 - |\rho_{g,n}^{11}(\theta)|)^{-3} \left( (|v_{g1,n}^{11}(\theta)| + |v_{g2,n}^{11}(\theta)| + C_7)^2 + \left( 1 - \Phi \left( (1 - |\rho_{g,n}^{11}(\theta)|)^{-1/2} \right) \right)^{-2} \right) \right) \right]^p \\ & \leq C_8 E \left[ \sup_{\theta \in \Theta} (|v_{g1,n}^{11}(\theta)| + |v_{g2,n}^{11}(\theta)| + C_7)^2 + C_9 \right]^p < \infty \end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , where the conclusion follows from Lemma E.3 in the same way as in (F.3). This concludes the proof that (F.7) and thus (F.5) are uniformly bounded. The SE property thus follows from Proposition 1 of Jenish and Prucha (2009).

PC: proof of  $\frac{1}{n} \sum_{g \in \mathcal{G}_n} (\sum_{a \in \mathcal{A}} [d_{g,n}^a f_{g,n}^a(\theta) - E[d_{g,n}^a f_{g,n}^a(\theta)]] ) \xrightarrow{P} 0$  as  $n \rightarrow \infty$  for  $\theta \in \Theta$

In order to establish the pointwise convergence, we apply Theorem D.1. As before, we will establish the result only for  $d_{g,n}^{11} f_{g,n}^{11}(\theta)$ ; the remaining terms can be analyzed analogously. We start by proving that the individual likelihood terms are  $L_1$ -NED on  $\{\eta_{g,n}\}_{g \in \mathcal{G}_n}$  with  $\eta_{g,n} = (X_{g,n}^s, X_{g,n}^o, u_{g,n}^s, u_{g,n}^o)$ . Note that

$$d_{g,n}^{11} f_{g,n}^{11}(\theta) = d_{g,n}^{11} \left( -\ln 2\pi - \frac{1}{2} (\ln |\Omega_{g,n}^{oo}(\theta)| + z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)) + \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) \right).$$

Given that  $\Omega_{g,n}^{oo}(\theta)$  is non-stochastic and its determinant is uniformly bounded away from zero by Lemma E.1, it suffices to establish the  $L_2$ -NED property for  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$ ,  $\{z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$ , and  $\{\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))\}_{g \in \mathcal{G}_n}$  and apply Theorem D.4 and Lemma D.10. In Lemma E.5, we have shown that  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field. Now we will apply Theorem D.2 to the remaining two random fields.

Let  $z_{g,n}^{(s)}(\theta) = E[z_{g,n}(\theta) | \mathcal{F}_{g,n}(s)]$ . Then by the elementwise mean value theorem, there exists  $\tilde{z}_{g,n}^{(s)}(\theta)$  with elements between  $z_{g,n}(\theta)$  and  $z_{g,n}^{(s)}(\theta)$  such that

$$|z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta) - z_{g,n}^{(s)'}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}^{(s)}(\theta)| \leq \|2\Omega_{g,n}^{oo-1}(\theta) \tilde{z}_{g,n}^{(s)}(\theta)\| \|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\|.$$

In order to verify conditions (ii) and (iii) of Theorem D.2, we have to show that  $\sup_s \|2\Omega_{g,n}^{oo-1}(\theta) \tilde{z}_{g,n}^{(s)}(\theta)\|_2$  and  $\sup_s \|2\Omega_{g,n}^{oo-1}(\theta) \tilde{z}_{g,n}^{(s)}(\theta) \|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\|_r$  are uniformly bounded, for some  $r > 2$ . Since the elements of  $\tilde{z}_{g,n}^{(s)}(\theta)$  lie between the elements of  $z_{g,n}(\theta)$  and  $z_{g,n}^{(s)}(\theta)$ , let  $C_{1g,n}^{(s)}(\theta)$  and  $C_{2g,n}^{(s)}(\theta)$  be  $2 \times 2$  diagonal matrices with elements in  $[0, 1]$  such that  $\tilde{z}_{g,n}^{(s)}(\theta) = C_{1g,n}^{(s)}(\theta) z_{g,n}(\theta) + C_{2g,n}^{(s)}(\theta) z_{g,n}^{(s)}(\theta)$ . Given that  $C_{1g,n}^{(s)}(\theta)$  and  $C_{2g,n}^{(s)}(\theta)$

have all elements in  $[0, 1]$  irrespectively of  $s, g, n$ , and  $\theta$ , it holds that

$$\begin{aligned}
\sup_s E \|2\Omega_{g,n}^{oo-1}(\theta) \tilde{z}_{g,n}^{(s)}(\theta)\|^2 &\leq 4\|\Omega_{g,n}^{oo-1}(\theta)\|^2 \sup_s E \|C_{1g,n}^{(s)}(\theta) z_{g,n}(\theta) + C_{2g,n}^{(s)}(\theta) z_{g,n}^{(s)}(\theta)\|^2 \\
&\leq C_3 \|\Omega_{g,n}^{oo-1}(\theta)\|^2 \left( E \|z_{g,n}(\theta)\|^2 + \sup_s E \|z_{g,n}^{(s)}(\theta)\|^2 \right) \\
&\leq 2C_3 \|\Omega_{g,n}^{oo-1}(\theta)\|^2 E \|z_{g,n}(\theta)\|^2 \\
&\leq 2C_3 \sup_{\theta \in \Theta} \|\Omega_{g,n}^{oo-1}(\theta)\|^2 E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^2 < \infty
\end{aligned}$$

for some constant  $C_3 > 0$  uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  by Lemmas E.1 and E.3, where the third inequality follows from the conditional Jensen's inequality. Next,

$$\begin{aligned}
&\sup_s E \|2\Omega_{g,n}^{oo-1}(\theta) \tilde{z}_{g,n}^{(s)}(\theta) \|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\|^r \\
&\leq 2^r \|\Omega_{g,n}^{oo-1}(\theta)\|^r \sup_s E \|(C_{1g,n}^{(s)}(\theta) z_{g,n}(\theta) + C_{2g,n}^{(s)}(\theta) z_{g,n}^{(s)}(\theta)) \|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\|^r \\
&\leq C_4 \|\Omega_{g,n}^{oo-1}(\theta)\|^r \sup_s E \left[ \|z_{g,n}(\theta)\| + \|z_{g,n}^{(s)}(\theta)\| \right]^{2r} \\
&\leq 2^{2r-1} C_4 \|\Omega_{g,n}^{oo-1}(\theta)\|^r \left( E \|z_{g,n}(\theta)\|^{2r} + \sup_s E \|z_{g,n}^{(s)}(\theta)\|^{2r} \right) \\
&\leq 2^{2r} C_4 \|\Omega_{g,n}^{oo-1}(\theta)\|^r E \|z_{g,n}(\theta)\|^{2r} \\
&\leq 2^{2r} C_4 \sup_{\theta \in \Theta} \|\Omega_{g,n}^{oo-1}(\theta)\|^r E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^{2r} < \infty
\end{aligned}$$

for some constant  $C_4 > 0$ . Since  $E \|z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)\|^2 \leq \sup_{\theta \in \Theta} \|\Omega_{g,n}^{oo-1}(\theta)\|^2 E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^4 < \infty$  uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  by Lemmas E.1 and E.3, condition (iv) of Theorem D.2 is fulfilled. As  $\{z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field by Lemma E.5, it follows that  $\{z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field as well.

Similarly, let  $v_{g,n}^{11(s)}(\theta) = E[v_{g,n}^{11}(\theta) | \mathcal{F}_{g,n}(s)]$  and verify the conditions of Theorem D.2 again. By the elementwise mean value theorem, there exists  $\tilde{v}_{g,n}^{11(s)}(\theta)$  between  $v_{g,n}^{11}(\theta)$  and  $v_{g,n}^{11(s)}(\theta)$  such that

$$\begin{aligned}
|\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) - \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))| &\leq \left| \frac{\partial \ln \Phi_2(\tilde{v}_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial v'} (v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)) \right| \\
&\leq C_5 \left( \|\tilde{v}_{g,n}^{11(s)}(\theta)\|^8 + C_6 \right) \|v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)\|
\end{aligned}$$

for some constants  $C_5, C_6 > 0$ , where the last inequality follows from Lemmas D.5, D.6, and E.2 by the

same argument as in (F.3). To bound  $\|\tilde{v}_{g,n}^{11(s)}(\theta)\|^8$  to verify condition (ii) of Theorem D.2, denote  $C_{7g,n}^{(s)}(\theta)$  and  $C_{8g,n}^{(s)}(\theta)$  the  $2 \times 2$  diagonal matrices with elements in  $[0, 1]$  such that  $\tilde{v}_{g,n}^{11(s)}(\theta) = C_{7g,n}^{(s)}(\theta)v_{g,n}^{11}(\theta) + C_{8g,n}^{(s)}(\theta)v_{g,n}^{11(s)}(\theta)$ . Since  $C_{7g,n}^{(s)}(\theta)$  and  $C_{8g,n}^{(s)}(\theta)$  have all elements in  $[0, 1]$  irrespectively of  $s, n, g$ , and  $\theta$ , it holds that

$$\begin{aligned} \sup_s E \|\tilde{v}_{g,n}^{11(s)}(\theta)\|^{16} &= \sup_s E \|C_{7g,n}^{(s)}(\theta)v_{g,n}^{11}(\theta) + C_{8g,n}^{(s)}(\theta)v_{g,n}^{11(s)}(\theta)\|^{16} \\ &\leq \sup_s C_9 \left( E \|v_{g,n}^{11}(\theta)\|^{16} + E \|v_{g,n}^{11(s)}(\theta)\|^{16} \right) \\ &\leq 2C_9 E \|v_{g,n}^{11}(\theta)\|^{16} \\ &\leq 2C_9 E \left[ \sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\| \right]^{16} \end{aligned}$$

is bounded uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  for some constant  $C_9 > 0$  by Lemma E.3. Next, condition (iii) of Theorem D.2 can be verified for some  $r > 2$  and some constant  $C_{10} > 0$  by

$$\begin{aligned} &\sup_s E \left[ \left( \|\tilde{v}_{g,n}^{11(s)}(\theta)\|^8 + C_6 \right) \|v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)\| \right]^r \\ &= \sup_s E \left[ \left( \|C_{7g,n}^{(s)}(\theta)v_{g,n}^{11}(\theta) + C_{8g,n}^{(s)}(\theta)v_{g,n}^{11(s)}(\theta)\|^8 + C_6 \right) \|v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)\| \right]^r \\ &\leq \sup_s E \left[ \left( C_{10} \left( \|v_{g,n}^{11}(\theta)\| + \|v_{g,n}^{11(s)}(\theta)\| \right)^8 + C_6 \right) \left( \|v_{g,n}^{11}(\theta)\| + \|v_{g,n}^{11(s)}(\theta)\| \right) \right]^r \\ &\leq \sup_s 2^{r-1} \left( C_{10}^r E \left[ \|v_{g,n}^{11}(\theta)\| + \|v_{g,n}^{11(s)}(\theta)\| \right]^{9r} + C_6^r E \left[ \|v_{g,n}^{11}(\theta)\| + \|v_{g,n}^{11(s)}(\theta)\| \right]^r \right) \\ &\leq \sup_s 2^{10r-2} C_{10}^r \left( E \|v_{g,n}^{11}(\theta)\|^{9r} + E \|v_{g,n}^{11(s)}(\theta)\|^{9r} \right) + \sup_s 2^{2r-2} C_6^r \left( E \|v_{g,n}^{11}(\theta)\|^r + E \|v_{g,n}^{11(s)}(\theta)\|^r \right) \\ &\leq 2^{10r-1} C_{10}^r E \|v_{g,n}^{11}(\theta)\|^{9r} + 2^{2r-1} C_6^r E \|v_{g,n}^{11}(\theta)\|^r \\ &\leq 2^{10r-1} C_{10}^r E \left[ \sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\| \right]^{9r} + 2^{2r-1} C_6^r E \left[ \sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\| \right]^r < \infty \end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  by Lemma E.3, where the second and third inequalities are implied by Loève's  $c_r$ -inequality, while the fourth inequality follows by the conditional Jensen's inequality. Finally, condition (iv) of Theorem D.2 can be verified in the same way as in (F.4) with the requirement that  $E \left[ \sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\| \right]^{18} < \infty$ , which is the case by Lemma E.3. As we have shown in Lemma E.5 that  $\{v_{g,n}^{11}(\theta)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field,  $\{\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field as well. Thus, by Theorem D.4 and Lemma D.10, it follows that  $\{d_{g,n}^{11} f_{g,n}^{11}(\theta)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_1$ -NED random field.

Hence, condition (i) of Theorem D.1 is satisfied, whereas condition (ii) is already verified in the beginning of the proof; condition (iii) is implied by Assumptions 2(i), 2(ii), and 4(i). Since convergence in probability

follows from convergence in  $L_1$ -norm, Theorem D.1 thus implies the pointwise convergence result.  $\square$

**Proof of Theorem 2:** By the elementwise mean value theorem, there exists  $\tilde{\theta}_n$  with elements between elements of  $\theta_0$  and  $\hat{\theta}_n$  such that

$$0 = \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} = \frac{\partial Q_n(\theta_0)}{\partial \theta} + \frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} (\hat{\theta}_n - \theta_0).$$

Once we show that  $\partial^2 Q_n(\tilde{\theta}_n)/\partial \theta \partial \theta' \xrightarrow{p} H(\theta_0)$  as  $n \rightarrow \infty$ , Assumption 12 will imply that, with a probability arbitrarily close to 1 as  $n \rightarrow \infty$ ,  $\partial^2 Q_n(\tilde{\theta}_n)/\partial \theta \partial \theta'$  is non-singular and it holds that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = - \left( \frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} \right)^{-1} \sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta}.$$

Therefore to prove the claim of the theorem, we will first establish that the term  $\sqrt{n} \partial Q_n(\theta_0)/\partial \theta$  converges in distribution to  $\mathcal{N}(0, J(\theta_0))$  as  $n \rightarrow \infty$ , and we will later show that  $\partial^2 Q_n(\tilde{\theta}_n)/\partial \theta \partial \theta' \xrightarrow{p} H(\theta_0)$  as  $n \rightarrow \infty$ .

Proof of  $\sqrt{n} \frac{\partial Q_n(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{g \in \mathcal{G}_n} \sum_{a \in \mathcal{A}} d_{g,n}^a \frac{\partial f_{g,n}^a(\theta_0)}{\partial \theta} \xrightarrow{d} \mathcal{N}(0, J(\theta_0))$  as  $n \rightarrow \infty$

---

We apply Theorem D.3. The individual score components have mean zero because the marginal likelihood contributions for each group are correctly specified. The remaining assumptions of Theorem D.3 concerning  $L_{2+\delta}$ -boundedness ( $\delta > 0$ ) and NED properties are verified at a general  $\theta \in \Theta$ , but they are applied at  $\theta = \theta_0$ .

By Loève's  $c_r$ -inequality and  $d_{g,n}^a$ ,  $a \in \mathcal{A}$ , being an indicator function, the individual score contributions are uniformly  $L_{2+\delta}$ -bounded if  $\sup_{n,g} E [\|\partial f_{g,n}^a(\theta)/\partial \theta\|]^{2+\delta} \leq \sup_{n,g} E [\sup_{\theta \in \Theta} \|\partial f_{g,n}^a(\theta)/\partial \theta\|]^{2+\delta} < \infty$  for some  $\delta > 0$ . The result for  $a = 11$  is a special case of the uniform boundedness of (F.5) verified in the proof of Theorem 1, property SE; the boundedness of the other terms can be proven in a similar way.

Now we establish that  $\{d_{g,n}^{11} \partial f_{g,n}^{11}(\theta)/\partial \theta\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field on the  $\alpha$ -mixing random field  $\{\eta_{g,n} = (X_{g,n}^s, X_{g,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g \in \mathcal{G}_n}$ . Recall that

$$d_{g,n}^{11} \frac{\partial f_{g,n}^{11}(\theta)}{\partial \theta} = -\frac{1}{2} d_{g,n}^{11} \frac{1}{|\Omega_{g,n}^{oo}(\theta)|} \frac{\partial |\Omega_{g,n}^{oo}(\theta)|}{\partial \theta} - \frac{1}{2} d_{g,n}^{11} \frac{\partial (z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta))}{\partial \theta} + d_{g,n}^{11} \frac{\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta}. \quad (\text{F.10})$$

By Theorem D.4, it suffices to show that each term of the summation is uniformly  $L_2$ -NED and to find their NED coefficients. We have already established in Lemma E.5 that  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{1/6}(s)$ , where  $\psi(s)$  is defined in Assumption 5. Since  $|\Omega_{g,n}^{oo}(\theta)|$  is uniformly

bounded away from zero and the norm of  $\partial|\Omega_{g,n}^{oo}(\theta)|/\partial\theta$  is uniformly bounded by Lemma E.1, the first term in (F.10) is uniformly  $L_2$ -NED with NED coefficients  $\psi^{1/6}(s)$ . For the second and third terms in (F.10), we apply Theorem D.2. Let  $d_{g,n}^{11(s)} = E[d_{g,n}^{11}|\mathcal{F}_{g,n}(s)]$  and  $z_{g,n}^{(s)}(\theta) = E[z_{g,n}(\theta)|\mathcal{F}_{g,n}(s)]$ . Then

$$\begin{aligned}
& \left\| d_{g,n}^{11} \frac{\partial(z'_{g,n}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))}{\partial\theta} - d_{g,n}^{11(s)} \frac{\partial(z_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}^{(s)}(\theta))}{\partial\theta} \right\| \\
& \leq \left( d_{g,n}^{11} + \left\| \frac{\partial(z_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}^{(s)}(\theta))}{\partial\theta} \right\| \right) \\
& \quad \times \left( |d_{g,n}^{11} - d_{g,n}^{11(s)}| + \left\| \frac{\partial(z'_{g,n}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))}{\partial\theta} - \frac{\partial(z_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}^{(s)}(\theta))}{\partial\theta} \right\| \right) \\
& \leq \left( 1 + \left\| \frac{\partial(z_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}^{(s)}(\theta))}{\partial\theta} \right\| \right) \\
& \quad \times \left( |d_{g,n}^{11} - d_{g,n}^{11(s)}| + \left\| \frac{\partial^2(\tilde{z}_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{z}_{g,n}^{(s)}(\theta))}{\partial\theta\partial z'} \right\| \|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\| \right) \\
& \leq \left( 1 + \left\| \frac{\partial(z_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}^{(s)}(\theta))}{\partial\theta} \right\| \right) \left( 1 + \left\| \frac{\partial^2(\tilde{z}_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{z}_{g,n}^{(s)}(\theta))}{\partial\theta\partial z'} \right\| \right) \\
& \quad \times \left( |d_{g,n}^{11} - d_{g,n}^{11(s)}| + \|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\| \right),
\end{aligned}$$

where the second inequality follows by the elementwise mean value theorem with elements of  $\tilde{z}_{g,n}^{(s)}(\theta)$  being between elements of  $z_{g,n}(\theta)$  and  $z_{g,n}^{(s)}(\theta)$ .

By the Cauchy-Schwartz, Minkowski's, and Liapunov's inequalities, conditions (ii) and (iii) of Theorem D.2 are fulfilled if  $\|\partial(z_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}^{(s)}(\theta))/\partial\theta\|_{4r}$ ,  $\|\partial^2(\tilde{z}_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{z}_{g,n}^{(s)}(\theta))/\partial\theta\partial z'\|_{4r}$ ,  $\|d_{g,n}^{11} - d_{g,n}^{11(s)}\|_{2r}$ , and  $\|z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)\|_{2r}$  are uniformly bounded for some  $r > 2$ . The boundedness of the first term can be proven in the same way as in (F.6) with an additional application of the conditional Jensen's inequality. Given the second order derivative of a quadratic form in Lemma D.7, it is not difficult to prove that  $\partial^2(\tilde{z}_{g,n}^{(s)'}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{z}_{g,n}^{(s)}(\theta))/\partial\theta\partial z'$  is uniformly  $L_{4r}$ -bounded. Trivially,  $d_{g,n}^{11} - d_{g,n}^{11(s)}$  is uniformly  $L_{2r}$ -bounded as well, while the  $L_{2r}$ -boundedness of  $z_{g,n}(\theta) - z_{g,n}^{(s)}(\theta)$  follows from Minkowski's and the conditional Jensen's inequalities and Lemma E.3. Since the uniform  $L_2$ -boundedness of  $\partial(z'_{g,n}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))/\partial\theta$  follows in the same way as in (F.6), condition (iv) of Theorem D.2 is fulfilled. Furthermore, since  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$  and  $\{z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$  are uniform  $L_2$ -NED random fields with NED coefficients  $\psi^{1/6}(s)$  by Lemma E.5, Theorem D.2 implies that the second term in (F.10) is uniformly  $L_2$ -NED with NED coefficients  $\psi^{(r-2)/(12r-12)}(s)$  for some  $r > 2$ .

Regarding the last term in (F.10), it follows similarly with  $v_{g,n}^{11(s)}(\theta) = E[v_{g,n}^{11}(\theta)|\mathcal{F}_{g,n}(s)]$  that

$$\begin{aligned} & \left\| d_{g,n}^{11} \frac{\partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta} - d_{g,n}^{11(s)} \frac{\partial \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta} \right\| \\ & \leq \left( 1 + \left\| \frac{\partial \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta} \right\| \right) \left( 1 + \left\| \frac{\partial^2 \ln \Phi_2(\tilde{v}_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta \partial v'} \right\| \right) (|d_{g,n}^{11} - d_{g,n}^{11(s)}| + \|v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)\|) \end{aligned}$$

with elements of  $\tilde{v}_{g,n}^{11(s)}(\theta)$  lying between elements of  $v_{g,n}^{11}(\theta)$  and  $v_{g,n}^{11(s)}(\theta)$ . Analogously to the previous case, applying Theorem D.2 requires us to check that  $\partial \ln \Phi_2(v_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))/\partial \theta$ ,  $\partial^2 \ln \Phi_2(\tilde{v}_{g,n}^{11(s)}(\theta), R_{g,n}^{11}(\theta))/\partial \theta \partial v'$  and  $d_{g,n}^{11} - d_{g,n}^{11(s)}$ ,  $v_{g,n}^{11}(\theta) - v_{g,n}^{11(s)}(\theta)$  are uniformly  $L_{4r}$ - and  $L_{2r}$ -bounded, respectively. Given Lemmas D.5 and D.6, the boundedness of the first term has been established in the proof of Theorem 1, property SE, and the boundedness of the second term can be established analogously. The third term is obviously uniformly  $L_{2r}$ -bounded, while the uniform  $L_{2r}$ -boundedness of the fourth term again follows from Minkowski's and the conditional Jensen's inequalities and Lemma E.3. Given that  $\{d_{g,n}\}_{g \in \mathcal{G}_n}$  and  $\{v_{g,n}^{11}(\theta)\}_{g \in \mathcal{G}_n}$  are uniform  $L_2$ -NED random fields with NED coefficients  $\psi^{1/6}(s)$  by Lemma E.5,  $\{d_{g,n}^{11} \partial \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))/\partial \theta\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{(r-2)/(12r-12)}(s)$  by Theorem D.2 for some  $r > 2$ . Further, it follows from Theorem D.4 that  $\{d_{g,n}^{11} \partial f_{g,n}^{11}(\theta)/\partial \theta\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{(r-2)/(12r-12)}(s)$ . Hence, conditions (iii) and (v) of Theorem D.3 are fulfilled, whereas conditions (iv) and (vi) are satisfied by Assumptions 10 and 2(i), 2(ii), and 9, respectively; condition (vii) is assumed in Assumption 12(ii). The asymptotic normality result thus follows from Theorem D.3.

Proof of  $\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} = \frac{1}{n} \sum_{g \in \mathcal{G}_n} \sum_{a \in \mathcal{A}} d_{g,n}^a \frac{\partial^2 f_{g,n}^a(\tilde{\theta}_n)}{\partial \theta \partial \theta'} \xrightarrow{p} H(\theta_0)$  as  $n \rightarrow \infty$

We can establish this result by showing that, for  $n \rightarrow \infty$ ,

$$\frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} - E \left[ \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} \right] \xrightarrow{p} 0 \text{ and } \frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{p} 0.$$

For the first claim, we apply Theorem D.1. As before, we establish the results for the part of the objective function term corresponding to index  $a = 11$  and a general  $\theta \in \Theta$  and apply it at  $\theta = \theta_0$ ; the results for the other terms can be proven in a similar way.



Note that

$$d_{g,n}^{11} \frac{\partial^2 f_{g,n}^{11}(\theta)}{\partial \theta \partial \theta'} = d_{g,n}^{11} \left[ -\frac{1}{2} \left( -\frac{1}{|\Omega_{g,n}^{oo}(\theta)|^2} \frac{\partial |\Omega_{g,n}^{oo}(\theta)|}{\partial \theta} \frac{\partial |\Omega_{g,n}^{oo}(\theta)|}{\partial \theta'} + \frac{1}{|\Omega_{g,n}^{oo}(\theta)|} \frac{\partial^2 |\Omega_{g,n}^{oo}(\theta)|}{\partial \theta \partial \theta'} \right. \right. \\ \left. \left. + \frac{\partial^2 (z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta))}{\partial \theta \partial \theta'} \right) + \frac{\partial^2 \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))}{\partial \theta \partial \theta'} \right].$$

We start by showing that  $d_{g,n}^{11} \partial^2 f_{g,n}^{11}(\theta) / \partial \theta \partial \theta'$  is uniformly  $L_p$ -bounded for some  $p > 1$ . Note that  $d_{g,n}^{11}$  is an indicator function, whereas  $|\Omega_{g,n}^{oo}(\theta)|$  is uniformly bounded away from zero and the norm of  $\partial |\Omega_{g,n}^{oo}(\theta)| / \partial \theta$  is uniformly bounded by Lemma E.1. It can be shown using the second order derivative of a determinant in Lemma D.7 that the norm of  $\partial^2 |\Omega_{g,n}^{oo}(\theta)| / \partial \theta \partial \theta'$  is uniformly bounded as well. Given the second order derivative of a quadratic form calculated in Lemma D.7, it is not difficult to see that the uniform  $L_p$ -boundedness of  $\partial^2 (z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)) / \partial \theta \partial \theta'$  can be easily established with the help of Lemmas E.1, E.3, and E.4 (analogously to the  $L_p$ -boundedness of the first derivative in (F.6)). Given the third result of Lemma D.5 and Lemma D.6, it can be also shown that  $\partial^2 \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) / \partial \theta \partial \theta'$  is  $L_p$ -bounded in a similar way as it was done for the first order derivative in the proof of Theorem 1, property SE.

In order to show that  $\{d_{g,n}^{11} \partial^2 f_{g,n}^{11}(\theta) / \partial \theta \partial \theta'\}_{g \in \mathcal{G}_n}$  is a uniform  $L_1$ -NED random field on the  $\alpha$ -mixing random field  $\{\eta_{g,n} = (X_{g,n}^s, X_{g,n}^o, u_{g,n}^s, u_{g,n}^o)\}_{g \in \mathcal{G}_n}$ , we have to establish the uniform  $L_2$ -NED property for  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$ , as is already done in Lemma E.5, and for the second order derivatives  $\{\partial^2 (z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)) / \partial \theta \partial \theta'\}_{g \in \mathcal{G}_n}$  and  $\{\partial^2 \ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta)) / \partial \theta \partial \theta'\}_{g \in \mathcal{G}_n}$  and apply Theorem D.4 and Lemma D.10. It can be done in a similar way as is done for  $\{z'_{g,n}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$  and  $\{\ln \Phi_2(v_{g,n}^{11}(\theta), R_{g,n}^{11}(\theta))\}_{g \in \mathcal{G}_n}$  in the proof of Theorem 1.

Finally, condition (iii) of Theorem D.1 is fulfilled by Assumptions 2(i), 2(ii), and 9. The convergence of the second order derivative of  $Q_n(\theta)$  at  $\theta_0$  to  $H(\theta_0)$  follows from Theorem D.1 and Assumption 12(i) and the fact that convergence in  $L_1$ -norm implies convergence in probability.

We continue by proving that, for  $n \rightarrow \infty$ ,

$$\frac{\partial^2 Q_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'} - \frac{\partial^2 Q_n(\theta_0)}{\partial \theta \partial \theta'} \xrightarrow{p} 0.$$

We apply the strategy used in the proof of Theorem 2 by Xu and Lee (2015a) and show that  $\{\partial^2 Q_n(\theta) / \partial \theta \partial \theta'\}_{g \in \mathcal{G}_n}$  is  $L_0$ -stochastically equicontinuous because the claim then follows directly from the proposition concerning the  $L_0$ -stochastic equicontinuity given in Andrews (1994). Since the objective function  $Q_n(\theta)$  as well as its second order derivative are continuously differentiable, the stochastic equicontinuity of  $\partial^2 Q_n(\theta) / \partial \theta \partial \theta'$  at

$\theta = \theta_0$  can however be established in a similar way as we have verified it for  $Q_n(\theta)$  in the proof of Theorem 1, property SE, which thus concludes the proof.  $\square$

## Acknowledgements

Authors thank Jaap Abbring, Raymond Florax, Ingmar Prucha, and Vaidotas Zemlys for valuable comments and discussions and Alfonso Flores-Lagunes for providing the code for the GMM estimator used in the Monte Carlo experiments. We also thank seminar participants at Tilburg University and VU Amsterdam and participants at the 14th International Workshop on Spatial Econometrics and Statistics (Paris, 2015), the 2nd Annual Conference of the International Association for Applied Econometrics (Thessaloniki, 2015), the 10th Meeting of the Netherlands Econometrics Study Group (Maastricht, 2015), the 4th Annual Lithuanian Conference on Economic Research (Kaunas, 2015), and the 26th  $(EC)^2$  Conference (Edinburgh, 2015) for helpful comments.

## References

- Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Andrews, D. (1994). Empirical process method in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*, pp. 2247–2294. Elsevier.
- Bai, Y., P. Song, and T. Raghunathan (2012). Joint composite estimating functions in spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74(5), 799–824.
- Booij, A. S., E. Leuven, and H. Oosterbeek (2015). Ability peer effects in university: Evidence from a randomized experiment. *DP No. 8769, IZA, Bonn, Germany*.
- Buchinsky, M. (1998). The dynamics of changes in the female wage distribution in the USA: A quantile regression approach. *Journal of Applied Econometrics* 13(1), 1–30.
- Calvò-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies* 76(4), 1239–1267.
- Chen, S. and Y. Zhou (2010). Semiparametric and nonparametric estimation of sample selection models under symmetry. *Journal of Econometrics* 157(1), 143–150.
- Conley, T. (1999). GMM estimation with cross sectional dependence. *Journal of Econometrics* 92(1), 1–45.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- Doğan, O. and S. Taşpınar (2014). Bayesian inference in sample selection models with spatial autoregressive processes. *Working paper*.
- Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review* 101(5), 1739–1774.
- Flores-Lagunes, A. and K. Schnier (2012). Estimation of sample selection models with spatial dependence. *Journal of Applied Econometrics* 27(2), 173–204.
- Greene, W. (2008). *Econometric Analysis*. Prentice-Hall.
- Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica* 42(4), 679–694.

- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47(1), 153–161.
- Horn, R. and C. Johnson (1985). *Matrix Analysis*. Cambridge University Press.
- Hoshino, T. (2016). Two-step estimation of incomplete information social interaction models with missing outcome data. *Working paper*, <http://ssrn.com/abstract=2675596>.
- Jenish, N. and I. R. Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150(1), 86–98.
- Jenish, N. and I. R. Prucha (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170(1), 178–190.
- Kelejian, H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* 17(1), 99–121.
- Kelejian, H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40(2), 509–533.
- Kelejian, H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics* 140(1), 131–154.
- Kelejian, H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157(1), 53–67.
- Kim, M. and Y. Sun (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics* 160(2), 349–371.
- Lee, L.-F. (2004). Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 76(6), 1899–1925.
- LeSage, J. and R. Pace (2009). *Introduction to Spatial Econometrics*. CRC Press.
- Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics* 28(4), 825–860.
- McMillen, D. (1995). Selection bias in spatial econometric models. *Journal of Regional Science* 35(3), 417–436.

- Mukherjee, B. and D. A. Singer (2010). International institutions and domestic compensation: The IMF and the politics of capital account liberalization. *American Journal of Political Science* 54(1), 45–60.
- Newey, W. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics, Vol. 4*, pp. 2111–2245. Elsevier.
- Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55(3), 703–708.
- Pinkse, J. and M. Slade (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics* 85(1), 125–154.
- Poirier, D. and P. Ruud (1988). Probit with dependent observations. *Review of Economic Studies* 55(4), 593–614.
- Qu, X. and L.-F. Lee (2012). LM tests for spatial correlation in spatial models with limited dependent variables. *Regional Science and Urban Economics* 42(3), 430–445.
- Robinson, P. (1982). On the asymptotic properties of estimators of models containing limited dependent variables. *Econometrica* 50(1), 27–41.
- Sharma, A., L. Siciliani, and A. Harris (2013). Waiting times and socioeconomic status: Does sample selection matter? *Economic Modelling* 33(0), 659–667.
- Su, L. and Z. Yang (2015). QML estimation of dynamic panel data models with spatial errors. *Journal of Econometrics* 185(1), 230–258.
- Vella, F. (1998). Estimating models with sample selection bias: A survey. *The Journal of Human Resources* 33(1), 127–169.
- Wang, H., E. M. Iglesias, and J. M. Wooldridge (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics* 172(1), 77–89.
- Wang, W. and L.-F. Lee (2013). Estimation of spatial autoregressive models with randomly missing data in the dependent variable. *The Econometrics Journal* 16(1), 73–102.
- Ward, P. S., R. J. G. M. Florax, and A. Flores-Lagunes (2014). Climate change and agricultural productivity in Sub-Saharan Africa: a spatial sample selection model. *European Review of Agricultural Economics* 41(2), 199–226.

- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.
- Xu, X. and L.-F. Lee (2015a). Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics* 188(1), 264–280.
- Xu, X. and L.-F. Lee (2015b). A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics* 186(1), 1–18.

# Supplementary Material for “Estimation of Spatial Sample Selection Models: A Partial Maximum Likelihood Approach”

Renata Rabovič\*      Pavel Čížek†

March 23, 2016

This supplementary material provides two additional technical lemmas and their proofs as well as proofs of Lemmas D.5–D.10 and E.1–E.5 in Appendices D and E, respectively.

## Appendix G    Some Additional Technical Lemmas

**Lemma G.1.** *Let  $X \sim \mathcal{N}(0, 1)$ . Then for any given  $r \in \mathbb{N}$ , there is some constant  $C_1 > 0$  such that for any  $c \in \mathbb{R}$ ,  $E[|X|^r | X \leq c] \leq |c|^{r-1} \phi(c)/\Phi(c) + (r-1)E[|X|^{r-2} | X \leq c]$  for  $r \geq 2$  with  $E[|X| | X \leq c] \leq \phi(c)/\Phi(c) + C_1$  and  $E[|X|^0 | X \leq c] = 1$ .*

*Proof.*

**Case 1.**  $c \leq 0$ :

Consider  $r = 0$ . Then  $E[|X|^0 | X \leq c] = \int_{-\infty}^c \phi(x)/\Phi(c) dx = \Phi(c)/\Phi(c) = 1$ . If  $r = 1$ , then  $E[|X| | X \leq c] = \int_{-\infty}^c (-x)\phi(x)/\Phi(c) dx = \int_{-\infty}^c \phi'(x)/\Phi(c) dx = \phi(c)/\Phi(c)$ , where the second equality follows by observing that  $\phi'(x) = -x\phi(x)$ . If  $r \geq 2$ , then by integration by parts,

$$\begin{aligned} E[|X|^r | X \leq c] &= \int_{-\infty}^c |x|^r \frac{\phi(x)}{\Phi(c)} dx = \int_{-\infty}^c (-x)^r \frac{\phi(x)}{\Phi(c)} dx = (-1)^{r-1} \int_{-\infty}^c x^{r-1} \frac{\phi'(x)}{\Phi(c)} dx \\ &= (-1)^{r-1} \left( c^{r-1} \frac{\phi(c)}{\Phi(c)} - (r-1) \int_{-\infty}^c x^{r-2} \frac{\phi(x)}{\Phi(c)} dx \right) \\ &= |c|^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) \int_{-\infty}^c (-x)^{r-2} \frac{\phi(x)}{\Phi(c)} dx = |c|^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1)E[|X|^{r-2} | X \leq c]. \end{aligned}$$

---

\*Corresponding author. CentER, Dept. of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands. **Email:** R.Rabovic@tilburguniversity.edu, Phone: +31 13 466 4412.

†CentER, Dept. of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000LE Tilburg, The Netherlands. **Email:** P.Cizek@tilburguniversity.edu, Phone: +31 13 466 8723.

**Case 2.**  $c > 0$ :

As in Case 1,  $E[|X|^0|X \leq c] = 1$ . If  $r = 1$ , then

$$\begin{aligned} E[|X||X \leq c] &= \int_{-\infty}^c |x| \frac{\phi(x)}{\Phi(c)} dx = \int_{-\infty}^0 (-x) \frac{\phi(x)}{\Phi(c)} dx + \int_0^c x \frac{\phi(x)}{\Phi(c)} dx = \int_{-\infty}^0 \frac{\phi'(x)}{\Phi(c)} dx - \int_0^c \frac{\phi'(x)}{\Phi(c)} dx \\ &= -\frac{\phi(c)}{\Phi(c)} + \frac{2\phi(0)}{\Phi(c)} \leq \frac{\phi(c)}{\Phi(c)} + \frac{2\phi(0)}{\Phi(0)} \leq \frac{\phi(c)}{\Phi(c)} + C_1. \end{aligned}$$

Consider  $r \geq 2$ . Then by integration by parts, it holds that

$$\begin{aligned} E[|X|^r|X \leq c] &= \int_{-\infty}^c |x|^r \frac{\phi(x)}{\Phi(c)} dx = \int_{-\infty}^0 (-x)^r \frac{\phi(x)}{\Phi(c)} dx + \int_0^c x^r \frac{\phi(x)}{\Phi(c)} dx \\ &= (-1)^{r-1} \int_{-\infty}^0 x^{r-1} \frac{\phi'(x)}{\Phi(c)} dx - \int_0^c x^{r-1} \frac{\phi'(x)}{\Phi(c)} dx \\ &= (-1)^{r-2} (r-1) \int_{-\infty}^0 x^{r-2} \frac{\phi(x)}{\Phi(c)} dx - c^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) \int_0^c x^{r-2} \frac{\phi(x)}{\Phi(c)} dx \\ &= -c^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) \int_{-\infty}^c |x|^{r-2} \frac{\phi(x)}{\Phi(c)} dx \leq |c|^{r-1} \frac{\phi(c)}{\Phi(c)} + (r-1) E[|X|^{r-2}|X \leq c]. \end{aligned}$$

□

**Lemma G.2.** Let  $Y_{g,n} = (Y_{gk,n})_{k=1}^K$  be a  $K$ -dimensional random vector. Then for some  $p \geq 1$ ,  $\{Y_{g,n}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_p$ -NED random field with NED coefficients  $\psi(s)$  if and only if for each  $k = 1, \dots, K$ ,  $\{Y_{gk,n}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_p$ -NED random field with NED coefficients  $\psi(s)$ .

*Proof.* We start with the ‘if’ part. By Loève’s  $c_r$ -inequality, it follows that

$$\begin{aligned} \|Y_{g,n} - E[Y_{g,n}|\mathcal{F}_{g,n}(s)]\|_p &= (E\|Y_{g,n} - E[Y_{g,n}|\mathcal{F}_{g,n}(s)]\|^p)^{1/p} \\ &= \left( E \left[ \sum_{k=1}^K |Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]|^2 \right]^{p/2} \right)^{1/p} \\ &\leq C_1 \left( \sum_{k=1}^K E|Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]|^p \right)^{1/p} \\ &\leq C_1 \sum_{k=1}^K \|Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]\|_p \\ &\leq \psi(s) C_1 \sum_{k=1}^K t_{g,n}^k, \end{aligned}$$

where  $\{t_{g,n}^k\}_{g \in \mathcal{G}_n}$  is the NED scaling factor for an element  $k$  with  $\sup_{n,g} \sum_{k=1}^K t_{g,n}^k < \infty$  because for each  $k = 1, \dots, K$ ,



$\{Y_{gk,n}\}_{g \in \mathcal{G}_n}$  is a uniform random field.

We continue with the ‘only if’ part:

$$\|Y_{gk,n} - E[Y_{gk,n}|\mathcal{F}_{g,n}(s)]\|_p \leq \|Y_{g,n} - E[Y_{g,n}|\mathcal{F}_{g,n}(s)]\|_p \leq t_{g,n}\psi(s),$$

where  $\{t_{g,n}\}_{g \in \mathcal{G}_n}$  is the NED scaling factor for random field  $\{Y_{g,n}\}_{g \in \mathcal{G}_n}$  with  $\sup_{n,g} t_{g,n} < \infty$  because  $\{Y_{g,n}\}_{g \in \mathcal{G}_n}$  is a uniform random field.  $\square$

## Appendix H Proof of Technical Lemmas in Appendix D

**Proof of Lemma D.5.** Noting that  $\partial \ln \Phi_2(v, R)/\partial v = \Phi_2(v, R)^{-1} \partial \Phi_2(v, R)/\partial v$  and  $\partial \ln \Phi_2(v, R)/\partial \theta = \Phi_2(v, R)^{-1} \partial \Phi_2(v, R)/\partial \theta$ , we apply differentiation under the integral sign twice to compute  $\partial \Phi_2(v, R)/\partial v$  and  $\partial \Phi_2(v, R)/\partial \theta$ :

$$\begin{aligned} \frac{\partial \Phi_2(v, R)}{\partial v} &= \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2 \frac{\partial v_1}{\partial v} + \int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R) dz_1 \frac{\partial v_2}{\partial v} \\ &= \left( \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2, \int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R) dz_1 \right)' \end{aligned} \quad (\text{H.1})$$

and

$$\frac{\partial \Phi_2(v, R)}{\partial \theta} = \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2 \frac{\partial v_1}{\partial \theta} + \int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R) dz_1 \frac{\partial v_2}{\partial \theta} + \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial \phi_2(z, R)}{\partial \theta} dz_2 dz_1. \quad (\text{H.2})$$

Note that

$$\phi_2((v_1, z_2)', R) = \frac{1}{\sqrt{1-\rho^2}} \phi(v_1) \phi\left(\frac{z_2 - \rho v_1}{\sqrt{1-\rho^2}}\right).$$

Thus,

$$\int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2 = \phi(v_1) \Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1-\rho^2}}\right). \quad (\text{H.3})$$

Similarly,

$$\int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R) dz_1 = \phi(v_2) \Phi\left(\frac{v_1 - \rho v_2}{\sqrt{1-\rho^2}}\right). \quad (\text{H.4})$$

The first claim now follows from (H.1), (H.3), and (H.4). Further, it is easy to show that

$$\frac{\partial \phi_2(z, R)}{\partial \theta} = -\frac{1}{2} \phi_2(z, R) \left( \frac{\partial \ln |R|}{\partial \theta} + \frac{\partial P(z, R)}{\partial \theta} \right). \quad (\text{H.5})$$

Hence,

$$\int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial \phi_2(z, R)}{\partial \theta} dz_2 dz_1 = -\frac{1}{2} \Phi_2(v, R) \left( \frac{\partial \ln |R|}{\partial \theta} + E_V \left[ \frac{\partial P(V, R)}{\partial \theta} \middle| V \leq v \right] \right), \quad (\text{H.6})$$

where  $V \sim \mathcal{N}(0, R)$ . The second conclusion follows by combining (H.2) with (H.3), (H.4), and (H.6). For the last claim of the lemma, note that

$$\frac{\partial^2 \ln \Phi_2(v, R)}{\partial \theta \partial \theta'} = \frac{1}{\Phi_2(v, R)} \frac{\partial^2 \Phi_2(v, R)}{\partial \theta \partial \theta'} - \frac{1}{\Phi_2^2(v, R)} \frac{\partial \Phi_2(v, R)}{\partial \theta} \frac{\partial \Phi_2(v, R)}{\partial \theta'}. \quad (\text{H.7})$$

Based on the results (H.2)–(H.4) and the definition of  $\xi(v, R)$ , it follows

$$\begin{aligned} \frac{1}{\Phi_2(v, R)} \frac{\partial^2 \Phi_2(v, R)}{\partial \theta \partial \theta'} &= \frac{1}{\Phi_2(v, R)} \frac{\partial v_1}{\partial \theta} \frac{\partial \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2}{\partial \theta'} + \xi_1(v, R) \frac{\partial^2 v_1}{\partial \theta \partial \theta'} \\ &\quad + \frac{1}{\Phi_2(v, R)} \frac{\partial v_2}{\partial \theta} \frac{\partial \int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R) dz_1}{\partial \theta'} + \xi_2(v, R) \frac{\partial^2 v_2}{\partial \theta \partial \theta'} \\ &\quad + \frac{1}{\Phi_2(v, R)} \frac{\partial \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial \phi_2(z, R)}{\partial \theta} dz_2 dz_1}{\partial \theta'}. \end{aligned} \quad (\text{H.8})$$

By applying differentiation under the integral sign, it follows as in (H.5)–(H.6) that

$$\begin{aligned} \frac{\partial \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2}{\partial \theta'} &= \phi_2(v, R) \frac{\partial v_2}{\partial \theta'} + \int_{-\infty}^{v_2} \frac{\partial \phi_2((v_1, z_2)', R)}{\partial \theta'} dz_2 \\ &= \phi_2(v, R) \frac{\partial v_2}{\partial \theta'} - \frac{1}{2} \phi(v_1) \Phi \left( \frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}} \right) \left( \frac{\partial \ln |R|}{\partial \theta'} + E_{\tilde{V}_2} \left[ \frac{\partial P((v_1, \tilde{V}_2)', R)}{\partial \theta'} \middle| \tilde{V}_2 \leq v_2 \right] \right). \end{aligned} \quad (\text{H.9})$$

Then by definition of  $\kappa(v, R)$ ,  $\xi(v, R)$ , and  $A(v, R)$ ,

$$\frac{1}{\Phi_2(v, R)} \frac{\partial v_1}{\partial \theta} \frac{\partial \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2}{\partial \theta'} = \kappa(v, R) \frac{\partial v_1}{\partial \theta} \frac{\partial v_2}{\partial \theta'} - \frac{1}{2} A(v, R). \quad (\text{H.10})$$

Symmetrically,

$$\frac{1}{\Phi_2(v, R)} \frac{\partial v_2}{\partial \theta} \frac{\partial \int_{-\infty}^{v_1} \phi_2((z_1, v_2)', R) dz_1}{\partial \theta'} = \kappa(v, R) \frac{\partial v_2}{\partial \theta} \frac{\partial v_1}{\partial \theta'} - \frac{1}{2} B(v, R). \quad (\text{H.11})$$

We proceed with the last term in (H.8). By applying differentiation under the integral sign twice,

$$\begin{aligned}
& \frac{1}{\Phi_2(v, R)} \frac{\partial \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial \phi_2(z, R)}{\partial \theta} dz_2 dz_1}{\partial \theta'} \\
&= \frac{1}{\Phi_2(v, R)} \left( \int_{-\infty}^{v_2} \frac{\partial \phi_2((v_1, z_2)', R)}{\partial \theta} dz_2 \frac{\partial v_1}{\partial \theta'} + \int_{-\infty}^{v_1} \frac{\partial \phi_2((z_1, v_2)', R)}{\partial \theta} dz_1 \frac{\partial v_2}{\partial \theta'} + \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'} dz_2 dz_1 \right) \\
&= -\frac{1}{2} \xi_1(v, R) \left( \frac{\partial \ln |R|}{\partial \theta} + E_{\tilde{V}_2} \left[ \frac{\partial P((v_1, \tilde{V}_2)', R)}{\partial \theta} \Big| \tilde{V}_2 \leq v_2 \right] \right) \frac{\partial v_1}{\partial \theta'} \\
&\quad - \frac{1}{2} \xi_2(v, R) \left( \frac{\partial \ln |R|}{\partial \theta} + E_{\tilde{V}_1} \left[ \frac{\partial P((\tilde{V}_1, v_2)', R)}{\partial \theta} \Big| \tilde{V}_1 \leq v_1 \right] \right) \frac{\partial v_2}{\partial \theta'} + \frac{1}{\Phi_2(v, R)} \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'} dz_2 dz_1 \\
&= -\frac{1}{2} A'(v, R) - \frac{1}{2} B'(v, R) + \frac{1}{\Phi_2(v, R)} \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'} dz_2 dz_1,
\end{aligned} \tag{H.12}$$

where the second equality follows in the same way as in (H.9). It can easily be shown that

$$\frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'} = -\frac{1}{2} \phi_2(z, R) G(z, R),$$

where  $G$  is defined (D.3). Thus,

$$\frac{1}{\Phi_2(v, R)} \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{\partial^2 \phi_2(z, R)}{\partial \theta \partial \theta'} dz_2 dz_1 = -\frac{1}{2} E_V [G(V, R) | V \leq v]. \tag{H.13}$$

The conclusion follows by combining (H.7) with (H.8), (H.10), (H.11), (H.12), and (H.13).  $\square$

**Proof of Lemma D.6.** We will start with the first claim by deriving the bounds when  $(v_1, v_2) \in (-1, +\infty) \times (-1, +\infty)$  and  $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$  and afterwards we will combine the results.

**Case 1.**  $(v_1, v_2) \in (-1, +\infty) \times (-1, +\infty)$ :

$$\frac{\phi(v_1) \Phi \left( \frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}} \right)}{\Phi_2(v, R)} \leq \frac{1}{\sqrt{2\pi} \Phi_2(-\iota_2, R)},$$

where  $\iota_2$  is a 2-dimensional vector of ones. Thus, we need to derive the lower bound for  $\Phi_2(-\iota_2, R)$ . Since  $R$  is a symmetric matrix, there exists an orthogonal matrix  $O$  such that  $R = O \text{Diag}\{\tau_1, \tau_2\} O'$ , where  $\tau_1 \leq \tau_2$  are the eigenvalues of  $R$ . Thus,  $R^{-1} = O \text{Diag}\{\tau_1^{-1}, \tau_2^{-1}\} O'$ . From Exercise 12.39 of Abadir and Magnus (2005), it holds

for any symmetric matrix  $A$  that  $z'Az \leq \max\text{eig}(A)z'z$ . Hence,

$$z'R^{-1}z = z'ODiag\{\tau_1^{-1}, \tau_2^{-1}\}O'z \leq \frac{1}{\tau_1}z'O O'z = \frac{1}{\tau_1}z'z = \left(\frac{z_1}{\sqrt{\tau_1}}\right)^2 + \left(\frac{z_2}{\sqrt{\tau_1}}\right)^2 \quad (\text{H.14})$$

and

$$\begin{aligned} \Phi_2(-\iota_2, R) &= \int_{-\infty}^{-1} \int_{-\infty}^{-1} \frac{1}{2\pi|R|^{1/2}} \exp\left(-\frac{1}{2}z'R^{-1}z\right) dz_2 dz_1 \\ &= \int_{-\infty}^{-1} \int_{-\infty}^{-1} \frac{1}{2\pi\sqrt{\tau_1\tau_2}} \exp\left(-\frac{1}{2}z'R^{-1}z\right) dz_2 dz_1 \\ &\geq \int_{-\infty}^{-1} \int_{-\infty}^{-1} \frac{1}{2\pi\sqrt{\tau_1\tau_2}} \exp\left(-\frac{1}{2}\left(\left(\frac{z_1}{\sqrt{\tau_1}}\right)^2 + \left(\frac{z_2}{\sqrt{\tau_1}}\right)^2\right)\right) dz_2 dz_1 \\ &= \sqrt{\frac{\tau_1}{\tau_2}} \int_{-\infty}^{-1} \int_{-\infty}^{-1} \frac{1}{\tau_1} \phi\left(\frac{z_1}{\sqrt{\tau_1}}\right) \phi\left(\frac{z_2}{\sqrt{\tau_1}}\right) dz_2 dz_1 \quad (\text{H.15}) \\ &= \sqrt{\frac{\tau_1}{\tau_2}} \Phi^2\left(\frac{-1}{\sqrt{\tau_1}}\right) \\ &\geq \sqrt{\frac{1-|\rho|}{2}} \Phi^2\left(\frac{-1}{\sqrt{1-|\rho|}}\right) \\ &= \left(\frac{1-|\rho|}{2}\right)^{1/2} \left(1 - \Phi\left((1-|\rho|)^{-1/2}\right)\right)^2, \end{aligned}$$

where the last inequality follows by noticing that  $\tau_1 = \min\{1-\rho, 1+\rho\} = 1-|\rho|$  and  $\tau_2 = \max\{1-\rho, 1+\rho\} < 2$ .

Hence,

$$\frac{\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)}{\Phi_2(v, R)} \leq C_1(1-|\rho|)^{-1/2} \left(1 - \Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2} \leq C_1(1-|\rho|)^{-7} \left(1 - \Phi\left((1-|\rho|)^{-1/2}\right)\right)^{-2}. \quad (\text{H.16})$$

**Case 2.**  $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$ :

First of all, we will derive the bound for  $\phi(v_1)\Phi\left(\frac{v_2-\rho v_1}{\sqrt{1-\rho^2}}\right)$ ; afterwards we will derive the bound for the entire

expression. Let  $\alpha = z^{\star'} R^{-1} z^{\star}$  with  $z^{\star} = (z_1^{\star}, z_2^{\star})' = \arg \min_z z' R^{-1} z$ , such that  $z \leq v$ . Then

$$\begin{aligned}
\phi(v_1) \Phi \left( \frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}} \right) &= \int_{-\infty}^{v_2} \phi_2((v_1, z_2)', R) dz_2 \\
&= \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \phi_2(z, R) \cdot \frac{\rho z_2 - z_1}{1 - \rho^2} dz_1 dz_2 \\
&= \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{\exp(-z' R^{-1} z/2)}{2\pi(1 - \rho^2)^{1/2}} \cdot \frac{\rho z_2 - z_1}{1 - \rho^2} dz_1 dz_2 \\
&\leq \frac{1}{2\pi(1 - \rho^2)^{3/2}} \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \exp \left( -\frac{1}{2} z' R^{-1} z \right) (|z_1| + |z_2|) dz_1 dz_2 \\
&\leq \frac{1}{2\pi(1 - \rho^2)^{3/2}} \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \exp \left( -\frac{1}{2} \alpha \right) \frac{\max\{6, \alpha\}^3}{(z' R^{-1} z)^3} (|z_1| + |z_2|) dz_1 dz_2 \\
&\leq \frac{\exp(-\alpha/2) \max\{6, \alpha\}^3}{2\pi(1 - |\rho|)^{3/2}} \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3} dz_1 dz_2,
\end{aligned}$$

where the second inequality follows from the following observation: the derivative of  $\exp(-\alpha/2)(\max\{6, \alpha\}/t)^3 / \exp(-t/2)$  indicates that the minimum of this function for  $t \geq \alpha$  is attained at  $t = \max\{6, \alpha\}$ ; the minimum of this function is at least 1. The double integral will be now proved to be bounded by a constant.

*Case (i).*  $v_1 \leq -1$  and  $v_2 \leq -1$ :

If  $z_1 \leq -1$  and  $z_2 \leq -1$ , then  $z' R^{-1} z = ((z_1 - z_2)^2 + 2(1 - \rho)z_1 z_2) / (1 - \rho^2) \geq 2(1 - \rho)z_1 z_2 / (1 - \rho^2) = 2z_1 z_2 / (1 + \rho) > z_1 z_2 > 0$ . Hence,

$$\begin{aligned}
\int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3} dz_1 dz_2 &\leq \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z_1 z_2)^3} dz_1 dz_2 = \int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \left( \frac{-1}{z_1^2 z_2^3} + \frac{-1}{z_1^3 z_2^2} \right) dz_1 dz_2 \\
&= \int_{-\infty}^{v_2} \left( \frac{1}{v_1 z_2^3} + \frac{1}{2v_1^2 z_2^2} \right) dz_2 = \frac{1}{2} \left( \frac{-1}{v_1 v_2^2} + \frac{-1}{v_1^2 v_2} \right) \leq 1.
\end{aligned}$$

*Case (ii).*  $-1 < v_1 \leq 1$  and  $v_2 \leq -1$ :

$$\int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3} dz_1 dz_2 = \int_{-\infty}^{v_2} \int_{-\infty}^{-1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3} dz_1 dz_2 + \int_{-\infty}^{v_2} \int_{-1}^{v_1} \frac{|z_1| + |z_2|}{(z' R^{-1} z)^3} dz_1 dz_2 \quad (\text{H.17})$$

The first double integral is bounded by a constant as it is shown in Case (i). Note that if  $-1 < z_1 \leq 1$  and  $z_2 \leq -1$ , then  $z'R^{-1}z = ((z_1 - \rho z_2)^2 + (1 - \rho^2)z_2^2)/(1 - \rho^2) \geq z_2^2 > 0$ . Thus,

$$\begin{aligned} \int_{-\infty}^{v_2} \int_{-1}^{v_1} \frac{|z_1| + |z_2|}{(z'R^{-1}z)^3} dz_1 dz_2 &\leq \int_{-\infty}^{v_2} \int_{-1}^{v_1} \frac{1 + |z_2|}{(z'R^{-1}z)^3} dz_1 dz_2 \leq \int_{-\infty}^{v_2} \int_{-1}^{v_1} \frac{1 + |z_2|}{z_2^6} dz_1 dz_2 \\ &= (v_1 + 1) \int_{-\infty}^{v_2} \left( \frac{1}{z_2^6} + \frac{-1}{z_2^5} \right) dz_2 \leq 2 \int_{-\infty}^{v_2} \left( \frac{1}{z_2^6} + \frac{-1}{z_2^5} \right) dz_2 \\ &= 2 \left( \frac{-1}{5v_2^5} + \frac{1}{4v_2^4} \right) \leq 2 \left( \frac{1}{5} + \frac{1}{4} \right) < 1. \end{aligned}$$

It concludes the proof that the integral in (H.17) is bounded by a constant.

*Case (iii).*  $v_1 > 1$  and  $v_2 \leq -1$ :

$$\int_{-\infty}^{v_2} \int_{-\infty}^{v_1} \frac{|z_1| + |z_2|}{(z'R^{-1}z)^3} dz_1 dz_2 = \int_{-\infty}^{v_2} \int_{-\infty}^1 \frac{|z_1| + |z_2|}{(z'R^{-1}z)^3} dz_1 dz_2 + \int_{-\infty}^{v_2} \int_1^{v_1} \frac{|z_1| + |z_2|}{(z'R^{-1}z)^3} dz_1 dz_2 \quad (\text{H.18})$$

We have already shown in Case (ii) that the first double integral is bounded by a constant. For the second integral, note that, if  $z_1 > 1$  and  $z_2 \leq -1$ , then  $z'R^{-1}z = ((z_1 + z_2)^2 - 2(1 + \rho)z_1 z_2)/(1 - \rho^2) \geq -2(1 + \rho)z_1 z_2/(1 - \rho^2) = -2z_1 z_2/(1 - \rho) > -z_1 z_2 > 0$ . Hence,

$$\begin{aligned} \int_{-\infty}^{v_2} \int_1^{v_1} \frac{|z_1| + |z_2|}{(z'R^{-1}z)^3} dz_1 dz_2 &\leq \int_{-\infty}^{v_2} \int_1^{v_1} \frac{|z_1| + |z_2|}{(-z_1 z_2)^3} dz_1 dz_2 = \int_{-\infty}^{v_2} \int_1^{v_1} \left( \frac{-1}{z_1^2 z_2^3} + \frac{1}{z_1^3 z_2^2} \right) dz_1 dz_2 \\ &= \int_{-\infty}^{v_2} \frac{1}{z_2^3} \left( \frac{1}{v_1} - 1 \right) + \frac{-1}{2z_2^2} \left( \frac{1}{v_1^2} - 1 \right) dz_2 = \frac{1}{2} \left( \frac{-1}{v_2^2} \left( \frac{1}{v_1} - 1 \right) + \frac{1}{v_2} \left( \frac{1}{v_1^2} - 1 \right) \right) \\ &\leq 1. \end{aligned}$$

It concludes the proof that the integral in (H.18) is bounded by a constant. Cases when  $v_1 \leq -1$  and  $-1 < v_2 \leq 1$  or  $v_2 > 1$  can be proven analogously. Thus for some constant  $C_4 > 0$ ,

$$\frac{\phi(v_1) \Phi \left( \frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}} \right)}{\Phi_2(v, R)} \leq \frac{C_4 (1 - |\rho|)^{-3/2} \exp(-\alpha/2) \max\{6, \alpha\}^3}{\Phi_2(v, R)}, \quad (\text{H.19})$$

if  $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$ . First, we will establish the bound for  $\exp(-\alpha/2)/\Phi_2(v, R)$ . The proof is similar to the proof of Proposition 3.2 of Hashorva and Hüsler (2003). Let  $t = (t_1, t_2)' = R^{-1}z^*$ . Then

$$\begin{aligned}
(z + z^*)' R^{-1}(z + z^*) &= z' R^{-1} z + 2z' R^{-1} z^* + z^{*\prime} R^{-1} z^* \\
&\leq \frac{1}{\tau_1} z' O O' z + 2z' R^{-1} z^* + z^{*\prime} R^{-1} z^* \\
&= \frac{1}{\tau_1} z' z + 2z' R^{-1} z^* + \alpha \\
&= \frac{1}{\tau_1} z' z + 2z' t + \alpha \\
&= \left( \frac{z_1}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_1 \right)^2 + \left( \frac{z_2}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_2 \right)^2 - (\sqrt{\tau_1} t_1)^2 - (\sqrt{\tau_1} t_2)^2 + \alpha,
\end{aligned}$$

where the inequality follows in the same way as in (H.14). Thus,

$$\begin{aligned}
\frac{\Phi_2(v, R)}{\exp(-\alpha/2)} &= \int_{-\infty}^{v_1} \int_{-\infty}^{v_2} \frac{1}{2\pi |R|^{1/2}} \exp\left(-\frac{1}{2}(z' R^{-1} z - \alpha)\right) dz_2 dz_1 \\
&= \int_{-\infty}^{v_1 - z_1^*} \int_{-\infty}^{v_2 - z_2^*} \frac{1}{2\pi \sqrt{\tau_1 \tau_2}} \exp\left(-\frac{1}{2}((z + z^*)' R^{-1}(z + z^*) - \alpha)\right) dz_2 dz_1 \\
&\geq \int_{-\infty}^{v_1 - z_1^*} \int_{-\infty}^{v_2 - z_2^*} \frac{1}{2\pi \sqrt{\tau_1 \tau_2}} \exp\left(-\frac{1}{2}\left(\left(\frac{z_1}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_1\right)^2 + \left(\frac{z_2}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_2\right)^2 - (\sqrt{\tau_1} t_1)^2 - (\sqrt{\tau_1} t_2)^2\right)\right) dz_2 dz_1 \\
&\geq \int_{-\infty}^0 \int_{-\infty}^0 \frac{1}{2\pi \sqrt{\tau_1 \tau_2}} \exp\left(-\frac{1}{2}\left(\left(\frac{z_1}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_1\right)^2 + \left(\frac{z_2}{\sqrt{\tau_1}} + \sqrt{\tau_1} t_2\right)^2 - (\sqrt{\tau_1} t_1)^2 - (\sqrt{\tau_1} t_2)^2\right)\right) dz_2 dz_1 \\
&= \frac{1}{2\pi} \sqrt{\frac{\tau_1}{\tau_2}} \frac{\Phi(\sqrt{\tau_1} t_1)}{\phi(\sqrt{\tau_1} t_1)} \frac{\Phi(\sqrt{\tau_1} t_2)}{\phi(\sqrt{\tau_1} t_2)} \\
&\geq \frac{1}{2\pi} \sqrt{\frac{1 - |\rho|}{2}} \frac{\Phi(\sqrt{\tau_1} t_1)}{\phi(\sqrt{\tau_1} t_1)} \frac{\Phi(\sqrt{\tau_1} t_2)}{\phi(\sqrt{\tau_1} t_2)}.
\end{aligned}$$

It follows from the proof of Lemma A.9 by Xu and Lee (2015) that  $\phi(x)/\Phi(x) \leq 2(|x| + C_2)$ . Thus,

$$\frac{\exp(-\alpha/2)}{\Phi_2(v, R)} \leq 8\sqrt{2}\pi(1 - |\rho|)^{-1/2} (|\sqrt{\tau_1} t_1| + C_2) (|\sqrt{\tau_1} t_2| + C_2) \leq C_5(1 - |\rho|)^{-1/2} (|t_1| + C_2) (|t_2| + C_2),$$

for some constant  $C_5 > 0$ , since  $|\tau_1| \leq 1$ .

It is not difficult to see that the solution to  $\min_z z' R^{-1} z$  s.t.  $z \leq v$  with  $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$  is unique and takes one of the three values  $(v_1, v_2)'$ ,  $(v_1, \rho v_1)'$ , or  $(\rho v_2, v_2)'$  depending on the values of  $v_1$ ,  $v_2$ , and  $\rho$  (similarly to Example 1 in Hashorva and Hüsler, 2003). If  $z^* = (v_1, v_2)'$ , then  $t = (v_1 - \rho v_2, v_2 - \rho v_1)' / (1 - \rho^2)$  and

$$\frac{\exp(-\alpha/2)}{\Phi_2(v, R)} \leq C_5(1 - |\rho|)^{-1/2} \left( \left| \frac{v_1 - \rho v_2}{1 - \rho^2} \right| + C_2 \right) \left( \left| \frac{v_2 - \rho v_1}{1 - \rho^2} \right| + C_2 \right) \leq C_5(1 - |\rho|)^{-5/2} (|v_1| + |v_2| + C_2)^2. \quad (\text{H.20})$$

If  $z^* = (v_1, \rho v_1)'$ , then  $t = (v_1, 0)'$  and

$$\frac{\exp(-\alpha/2)}{\Phi_2(v, R)} \leq C_5(1 - |\rho|)^{-1/2}(|v_1| + C_2)C_2 \leq C_5(1 - |\rho|)^{-5/2}(|v_1| + |v_2| + C_2)^2. \quad (\text{H.21})$$

The bound when  $z^* = (\rho v_2, v_2)'$  can be derived analogously. Next, we calculate the bound for  $\alpha^3$ :

$$\begin{aligned} \alpha^3 &\leq (1 - \rho^2)^{-3}(v_1^2 - 2\rho v_1 v_2 + v_2^2)^3 \leq (1 - \rho^2)^{-3}(|v_1|^2 + 2|v_1||v_2| + |v_2|^2)^3 \\ &= (1 - \rho^2)^{-3}(|v_1| + |v_2|)^6 \leq (1 - |\rho|)^{-3}(|v_1| + |v_2|)^6. \end{aligned}$$

Hence,

$$\max\{6, \alpha\}^3 \leq (1 - |\rho|)^{-3}(|v_1| + |v_2| + C_2)^6. \quad (\text{H.22})$$

It follows from combining (H.19), (H.20), (H.21), and (H.22) that

$$\frac{\phi(v_1)\Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} \leq C_1(1 - |\rho|)^{-7}(|v_1| + |v_2| + C_2)^8, \quad (\text{H.23})$$

if  $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$ . The conclusion is obtained by combining (H.16) with (H.23).

We continue with the second claim of the lemma. If  $(v_1, v_2) \in (-1, +\infty) \times (-1, +\infty)$ , then clearly

$$\begin{aligned} \frac{\phi_2(v, R)}{\Phi_2(v, R)} &\leq \frac{1}{2\pi(1 - \rho^2)^{1/2}\Phi_2(-\iota_2, R)} \leq \frac{1}{2\pi(1 - |\rho|)^{1/2}\Phi_2(-\iota_2, R)} \leq C_1(1 - |\rho|)^{-1} \left(1 - \Phi\left((1 - |\rho|)^{-1/2}\right)\right)^{-2} \\ &\leq C_1(1 - |\rho|)^{-3} \left(1 - \Phi\left((1 - |\rho|)^{-1/2}\right)\right)^{-2}, \end{aligned} \quad (\text{H.24})$$

where the second inequality follows from (H.15). If  $(v_1, v_2) \notin (-1, +\infty) \times (-1, +\infty)$ ,

$$\frac{\phi_2(v, R)}{\Phi_2(v, R)} = \frac{\exp(-v'R^{-1}v/2)}{2\pi(1 - \rho^2)^{1/2}\Phi_2(v, R)} \leq \frac{\exp(-\alpha/2)}{2\pi(1 - |\rho|)^{1/2}\Phi_2(v, R)} \leq C_1(1 - |\rho|)^{-3}(|v_1| + |v_2| + C_2)^2, \quad (\text{H.25})$$

where the result follows from (H.20) and (H.21). The conclusion is obtained by combining (H.24) and (H.25).

The third claim follows in the same way as in (H.15) with  $-\iota_2$  replaced with the 2-dimensional vector of zeros.

Thus,

$$\Phi_2(0, R) \geq \frac{(1 - |\rho|)^{1/2}}{2^{5/2}} \geq C_3(1 - |\rho|)^{1/2}.$$

□

**Proof of Lemma D.7.** Given a matrix function  $F$  and a matrix  $X$ , we proceed as follows: (i) compute the differ-



ential of  $F(X)$ , (ii) vectorize to obtain  $d \text{vec } F(X) = A(X) d \text{vec } X$ , and (iii) conclude that  $\partial \text{vec } F(X) / \partial (\text{vec } X)' = A(X)$  (see Magnus and Neudecker, 1999, for more details). The differential of the first function is given by  $d|F(\theta)| = |F(\theta)| \text{Tr}(F^{-1}(\theta) dF(\theta)) = |F(\theta)| (\text{vec } F^{-1}(\theta))' d \text{vec } F(\theta) = |F(\theta)| (\text{vec } F^{-1}(\theta))' (\partial \text{vec } F(\theta) / \partial \theta') d\theta$ , where we used that  $\text{Tr}(A'B) = (\text{vec } A)' \text{vec } B$  and  $F(\theta)$  is symmetric implying that  $F^{-1}(\theta)$  is symmetric as well. Hence,

$$\frac{\partial |F(\theta)|}{\partial \theta'} = |F(\theta)| (\text{vec } F^{-1}(\theta))' \frac{\partial \text{vec } F(\theta)}{\partial \theta'}.$$

Thus given the definition of  $K(\theta)$ ,

$$\frac{\partial |F(\theta)|}{\partial \theta} = |F(\theta)| K(\theta) \text{vec } F^{-1}(\theta).$$

In order to obtain the second order derivative, we calculate the differential once more:

$$\begin{aligned} d \left( \frac{\partial |F(\theta)|}{\partial \theta} \right) &= d|F(\theta)| K(\theta) \text{vec } F^{-1}(\theta) + |F(\theta)| dK(\theta) \text{vec } F^{-1}(\theta) + |F(\theta)| K(\theta) d \text{vec } F^{-1}(\theta) \\ &= K(\theta) \text{vec } F^{-1}(\theta) d|F(\theta)| + |F(\theta)| ((\text{vec } F^{-1}(\theta))' \otimes I_p) d \text{vec } K(\theta) + |F(\theta)| K(\theta) d \text{vec } F^{-1}(\theta) \\ &= \left( K(\theta) \text{vec } F^{-1}(\theta) \frac{\partial |F(\theta)|}{\partial \theta'} + |F(\theta)| ((\text{vec } F^{-1}(\theta))' \otimes I_p) \frac{\partial \text{vec } K(\theta)}{\partial \theta'} + |F(\theta)| K(\theta) \frac{\partial \text{vec } F^{-1}(\theta)}{\partial \theta'} \right) d\theta, \end{aligned}$$

where the second equality follows from  $\text{vec}(ABC) = (C' \otimes A) \text{vec } B$ . The result follows.

Next,

$$\begin{aligned} d(f'(\theta) F^{-1}(\theta) f(\theta)) &= 2f'(\theta) F^{-1}(\theta) df(\theta) + f'(\theta) dF^{-1}(\theta) f(\theta) \\ &= 2f'(\theta) F^{-1}(\theta) df(\theta) + (f'(\theta) \otimes f'(\theta)) d \text{vec } F^{-1}(\theta) \\ &= (2f'(\theta) F^{-1}(\theta) \partial f(\theta) / \partial \theta' + (f'(\theta) \otimes f'(\theta)) \partial \text{vec } F^{-1}(\theta) / \partial \theta') d\theta, \end{aligned}$$

where the second equality follows from  $\text{vec}(ABC) = (C' \otimes A) \text{vec } B$ . Thus,

$$\frac{\partial (f'(\theta) F^{-1}(\theta) f(\theta))}{\partial \theta'} = 2f'(\theta) F^{-1}(\theta) \frac{\partial f(\theta)}{\partial \theta'} + (f'(\theta) \otimes f'(\theta)) \frac{\partial \text{vec } F^{-1}(\theta)}{\partial \theta'}.$$

The result is obtained by taking a transpose of this expression. We continue with the second differential:

$$\begin{aligned}
d\left(\frac{\partial(f'(\theta)F^{-1}(\theta)f(\theta))}{\partial\theta}\right) &= 2dL(\theta)F^{-1}(\theta)f(\theta) + 2L(\theta)dF^{-1}(\theta)f(\theta) + 2L(\theta)F^{-1}(\theta)df(\theta) \\
&\quad + dM(\theta)(f(\theta) \otimes f(\theta)) + M(\theta)d(f(\theta) \otimes f(\theta)) \\
&= 2(f'(\theta)F^{-1}(\theta) \otimes I_p) d\text{vec } L(\theta) + 2(f'(\theta) \otimes L(\theta)) d\text{vec } F^{-1}(\theta) + 2L(\theta)F^{-1}(\theta)df(\theta) \\
&\quad + (f'(\theta) \otimes f'(\theta) \otimes I_p) d\text{vec } M(\theta) + M(\theta)(K_{1n} \otimes I_n) [(I_n \otimes f(\theta))df(\theta) + (f(\theta) \otimes I_n)df(\theta)] \\
&= 2(f'(\theta)F^{-1}(\theta) \otimes I_p) d\text{vec } L(\theta) + 2(f'(\theta) \otimes L(\theta)) d\text{vec } F^{-1}(\theta) + (f'(\theta) \otimes f'(\theta) \otimes I_p) d\text{vec } M(\theta) \\
&\quad + (2L(\theta)F^{-1}(\theta) + M(\theta)(K_{1n} \otimes I_n) [(I_n \otimes f(\theta)) + (f(\theta) \otimes I_n)]) df(\theta) \\
&= \left(2(f'(\theta)F^{-1}(\theta) \otimes I_p) \frac{\partial \text{vec } L(\theta)}{\partial\theta'} + 2(f'(\theta) \otimes L(\theta)) \frac{\partial \text{vec } F^{-1}(\theta)}{\partial\theta'} + (f'(\theta) \otimes f'(\theta) \otimes I_p) \frac{\partial \text{vec } M(\theta)}{\partial\theta'} \right. \\
&\quad \left. + (2L(\theta)F^{-1}(\theta) + M(\theta)(K_{1n} \otimes I_n) [(I_n \otimes f(\theta)) + (f(\theta) \otimes I_n)]) \frac{\partial f(\theta)}{\partial\theta'}\right) d\theta,
\end{aligned}$$

where the second equality follows from  $\text{vec}(ABC) = (C' \otimes A) \text{vec } B$  and for  $X$  and  $Y$  being  $n \times q$  and  $p \times r$  matrices,  $d\text{vec}(X \otimes Y) = (I_q \otimes K_{rn} \otimes I_p) [(I_{nq} \otimes \text{vec } Y) d\text{vec } X + (\text{vec } X \otimes I_{pr}) d\text{vec } Y]$  as derived in Magnus and Neudecker (1999, p. 185). The conclusion follows.  $\square$

**Proof of Lemma D.8.**

$$\begin{aligned}
\|A \otimes B\| &= \sqrt{\sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^p \sum_{k=1}^q (A_{ij} B_{lk})^2} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \sum_{l=1}^p \sum_{k=1}^q B_{lk}^2} \\
&= \sqrt{\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \|B\|^2} = \sqrt{\|A\|^2 \|B\|^2} = \|A\| \|B\|.
\end{aligned}$$

$\square$

**Proof of Lemma D.9.** Based on equation (9) in Muthén (1990),

$$\begin{aligned}
E[X_1^2 | X \leq v] &= 1 - v_1 \frac{\phi(v_1) \Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} - \rho^2 v_2 \frac{\phi(v_2) \Phi\left(\frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} + \rho \sqrt{1 - \rho^2} \frac{\phi(v_2) \phi\left(\frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} \\
&= 1 - v_1 \frac{\phi(v_1) \Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} - \rho^2 v_2 \frac{\phi(v_2) \Phi\left(\frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} + \rho(1 - \rho^2) \frac{\phi_2(v, R)}{\Phi_2(v, R)} \\
&= 1 - v_1 \xi_1(v, R) - \rho^2 v_2 \xi_2(v, R) + \rho(1 - \rho^2) \kappa(v, R),
\end{aligned}$$

where the second equality follows by observing that  $\phi(v_2) \phi((v_1 - \rho v_2)/\sqrt{1 - \rho^2})/\sqrt{1 - \rho^2} = \phi_2(v, R)$ , whereas the

last equality follows by the definitions of  $\xi(v, R)$  and  $\kappa(v, R)$  in (D.1) and (D.2), respectively. Symmetrically,

$$E[X_2^2|X \leq v] = 1 - v_2\xi_2(v, R) - \rho^2 v_1\xi_1(v, R) + \rho(1 - \rho^2)\kappa(v, R).$$

From equation (11) in Muthén (1990),

$$\begin{aligned} E[X_1 X_2|X \leq v] &= \rho - \rho v_1 \frac{\phi(v_1)\Phi\left(\frac{v_2 - \rho v_1}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} - \rho v_2 \frac{\phi(v_2)\Phi\left(\frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} + \sqrt{1 - \rho^2} \frac{\phi(v_2)\phi\left(\frac{v_1 - \rho v_2}{\sqrt{1 - \rho^2}}\right)}{\Phi_2(v, R)} \\ &= \rho - \rho v_1 \xi_1(v, R) - \rho v_2 \xi_2(v, R) + (1 - \rho^2)\kappa(v, R). \end{aligned}$$

The conclusion follows by noticing that  $E[XX'|X \leq v] = (E[X_1^2|X \leq v] \ E[X_1 X_2|X \leq v]; \ E[X_1 X_2|X \leq v] \ E[X_2^2|X \leq v])$ .  $\square$

**Proof of Lemma D.10.** The proof closely follows the proof of Theorem 17.9 of Davidson (1994). Let  $X_{i,n}^{(s)} = E[X_{i,n}|\mathcal{F}_{i,n}(s)]$  and  $Y_{i,n}^{(s)} = E[Y_{i,n}|\mathcal{F}_{i,n}(s)]$ . Then

$$\begin{aligned} &\|X_{i,n}Y_{i,n} - E[X_{i,n}Y_{i,n}|\mathcal{F}_{i,n}(s)]\|_p \\ &= \|(X_{i,n}Y_{i,n} - X_{i,n}Y_{i,n}^{(s)}) + (X_{i,n}Y_{i,n}^{(s)} - X_{i,n}^{(s)}Y_{i,n}^{(s)}) - E[(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})|\mathcal{F}_{i,n}(s)]\|_p \\ &\leq \|X_{i,n}(Y_{i,n} - Y_{i,n}^{(s)})\|_p + \|Y_{i,n}^{(s)}(X_{i,n} - X_{i,n}^{(s)})\|_p + \|E[(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})|\mathcal{F}_{i,n}(s)]\|_p \\ &\leq \|X_{i,n}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} + \|Y_{i,n}^{(s)}\|_{2p}\|X_{i,n} - X_{i,n}^{(s)}\|_{2p} + \|E[(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})|\mathcal{F}_{i,n}(s)]\|_p \\ &\leq \|X_{i,n}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} + \|Y_{i,n}\|_{2p}\|X_{i,n} - X_{i,n}^{(s)}\|_{2p} + \|(X_{i,n} - X_{i,n}^{(s)})(Y_{i,n} - Y_{i,n}^{(s)})\|_p \\ &\leq \|X_{i,n}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} + \|Y_{i,n}\|_{2p}\|X_{i,n} - X_{i,n}^{(s)}\|_{2p} + \|X_{i,n} - X_{i,n}^{(s)}\|_{2p}\|Y_{i,n} - Y_{i,n}^{(s)}\|_{2p} \\ &\leq \|X_{i,n}\|_{2p}t_{i,n}^Y\psi^Y(s) + \|Y_{i,n}\|_{2p}t_{i,n}^X\psi^X(s) + t_{i,n}^X\psi^X(s)t_{i,n}^Y\psi^Y(s) \\ &\leq t_{i,n}\psi(s), \end{aligned}$$

where the first and second inequalities are implied by the Minkowski's and Cauchy-Schwartz inequalities, respectively, whereas the third inequality follows by the conditional Jensen's inequality and law of iterated expectations; the fourth inequality again follows by the Cauchy-Schwartz inequality. The final claim of the lemma follows from Definition 2.  $\square$

## Appendix I Proof of Lemmas in Appendix E

**Proof of Lemma E.1.** (i) Let  $\tau_{1g,n}^b(\theta) \leq \tau_{2g,n}^b(\theta)$  be the eigenvalues of  $\Omega_{g,n}^b(\theta)$ . Then

$$\inf_{n,g} \inf_{\theta \in \Theta} |\Omega_{g,n}^b(\theta)| = \inf_{n,g} \inf_{\theta \in \Theta} (\tau_{1g,n}^b(\theta) \tau_{2g,n}^b(\theta)) \geq \inf_{n,g} \inf_{\theta \in \Theta} \tau_{1g,n}^{b^2}(\theta) > 0$$

by Assumption 6.

In the same way as in the proof of Lemma 2 by Xu and Lee (2015), it follows that  $\inf_{n,i} \inf_{\theta \in \Theta} \Omega_{ii,n}^b(\theta) \geq \inf_n (\|I_{2n} - \lambda^b W_n^b\|_\infty \|I_{2n} - \lambda^b W_n^b\|_1 \cdot \min\{1, \sigma^2\})^{-1} > 0$  by Assumptions 1(ii), 2(i), and 7.

(ii) Next, let  $d, e \in \{s, o\}$ . Then uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$ ,  $\|\Omega_{g,n}^c(\theta)\| \leq C_1 \|\Omega_{g,n}^c(\theta)\|_\infty \leq C_1 \|\Omega_n^c(\theta)\|_\infty \leq C_1 C_2 \|(I_{2n} - \lambda^d W_n^d)^{-1} (I_{2n} - \lambda^e W_n^{e'})^{-1}\|_\infty \leq C_1 C_2 \|(I_{2n} - \lambda^d W_n^d)^{-1}\|_\infty \|(I_{2n} - \lambda^e W_n^e)^{-1}\|_1 < \infty$  for some constants  $C_1, C_2 > 0$ . The first inequality is implied by the equivalence of matrix norms on finite dimensional matrix spaces, whereas the third inequality follows by compactness of the parameter space. The conclusion is implied by Assumption 1(ii).

Next, note that

$$\left\| \frac{\partial \text{vec } \Omega_{g,n}^c(\theta)}{\partial \theta'} \right\| = \sqrt{\left\| \frac{\partial \Omega_{g,n}^c(\theta)}{\partial \lambda^s} \right\|^2 + \left\| \frac{\partial \Omega_{g,n}^c(\theta)}{\partial \lambda^o} \right\|^2 + \left\| \frac{\partial \Omega_{g,n}^c(\theta)}{\partial \rho} \right\|^2 + \left\| \frac{\partial \Omega_{g,n}^c(\theta)}{\partial \sigma^2} \right\|^2}.$$

We will show that  $\|\partial \Omega_{g,n}^c(\theta) / \partial \lambda^s\|$  is uniformly bounded, while the boundedness of the other terms can be established in a similar way. For some constant  $C_3 > 0$ , uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$

$$\begin{aligned} \left\| \frac{\partial \Omega_{g,n}^c(\theta)}{\partial \lambda^s} \right\| &\leq C_3 \left\| \frac{\partial \Omega_{g,n}^c(\theta)}{\partial \lambda^s} \right\|_\infty \\ &\leq C_3 \left\| \frac{\partial \Omega_n^c(\theta)}{\partial \lambda^s} \right\|_\infty \\ &\leq C_2 C_3 \left\| \frac{\partial ((I_{2n} - \lambda^d W_n^d)^{-1} (I_{2n} - \lambda^e W_n^{e'})^{-1})}{\partial \lambda^s} \right\|_\infty \\ &\leq C_2 C_3 \left( \|(I_{2n} - \lambda^d W_n^d)^{-1} W_n^d (I_{2n} - \lambda^d W_n^d)^{-1} (I_{2n} - \lambda^e W_n^{e'})^{-1}\|_\infty \right. \\ &\quad \left. + \|(I_{2n} - \lambda^d W_n^d)^{-1} (I_{2n} - \lambda^e W_n^{e'})^{-1} W_n^{e'} (I_{2n} - \lambda^e W_n^{e'})^{-1}\|_\infty \right) < \infty. \end{aligned}$$

The first inequality follows by the equivalence of matrix norms on finite dimensional matrix spaces. The result is implied by sub-multiplicity of matrix norms and Assumption 1(ii).

(iii) Note that  $\|\Omega_{g,n}^{b-1}(\theta)\| = |\Omega_{g,n}^b(\theta)|^{-1} \|\Omega_{g,n}^b(\theta)\| < \infty$  uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$  by parts (i) and

(ii).

Consequently, by Lemma D.8,

$$\begin{aligned} \left\| \frac{\partial \text{vec } \Omega_{g,n}^{b-1}(\theta)}{\partial \theta'} \right\| &= \left\| (\Omega_{g,n}^{b-1}(\theta) \otimes \Omega_{g,n}^{b-1}(\theta)) \frac{\partial \text{vec } \Omega_{g,n}^b(\theta)}{\partial \theta'} \right\| \\ &\leq \|\Omega_{g,n}^{b-1}(\theta)\|^2 \left\| \frac{\partial \text{vec } \Omega_{g,n}^b(\theta)}{\partial \theta'} \right\| < \infty. \end{aligned}$$

Lemma D.7 implies that

$$\begin{aligned} \left\| \frac{\partial |\Omega_{g,n}^b(\theta)|}{\partial \theta} \right\| &= \left\| |\Omega_{g,n}^b(\theta)| \left( \frac{\partial \text{vec } \Omega_{g,n}^b(\theta)}{\partial \theta'} \right)' \text{vec } \Omega_{g,n}^{b-1}(\theta) \right\| \\ &\leq |\Omega_{g,n}^b(\theta)| \left\| \frac{\partial \text{vec } \Omega_{g,n}^b(\theta)}{\partial \theta'} \right\| \|\Omega_{g,n}^{b-1}(\theta)\| < \infty \end{aligned}$$

uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$ , where the boundedness of  $|\Omega_{g,n}^b(\theta)|$  is implied by the boundedness of  $\|\Omega_{g,n}^b(\theta)\|$ .  $\square$

**Proof of Lemma E.2.** (i) From Exercise 12.39 in Abadir and Magnus (2005), for any symmetric matrix  $A$  and compatible vector  $x$ ,  $x'Ax \geq \text{mineig}(A)x'x$ , where  $\text{mineig}(A)$  is the minimum eigenvalue of  $A$ . Let  $\tau_{1g,n}^{11}(\theta) \leq \tau_{2g,n}^{11}(\theta)$  be the eigenvalues of  $\Sigma_{g,n}^{11}(\theta)$  and consider vectors  $x = (1, 0)'$  and  $x = (0, 1)'$  for  $j = 1$  and  $j = 2$ , respectively. Then  $\inf_{n,g} \inf_{\theta \in \Theta} \Sigma_{gjj,n}^{11}(\theta) \geq \inf_{n,g} \inf_{\theta \in \Theta} \tau_{1g,n}^{11}(\theta) > 0$  by Assumption 6.

Next, note that

$$|\Sigma_{g,n}^{11}(\theta)| = \Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta) - \Sigma_{g12,n}^{11}(\theta) = \Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)(1 - \rho_{g,n}^{112}(\theta)).$$

Thus,

$$|\rho_{g,n}^{11}(\theta)| = \sqrt{1 - \frac{|\Sigma_{g,n}^{11}(\theta)|}{\Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)}} = \sqrt{1 - \frac{\tau_{1g,n}^{11}(\theta)\tau_{2g,n}^{11}(\theta)}{\Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)}} \leq \sqrt{1 - \frac{\tau_{1g,n}^{112}(\theta)}{\Sigma_{g11,n}^{11}(\theta)\Sigma_{g22,n}^{11}(\theta)}} < 1$$

uniformly in  $g \in \mathcal{G}_n$ ,  $n \in \mathbb{N}$ , and  $\theta \in \Theta$  by Assumption 6 and the fact that  $\inf_{n,g} \inf_{\theta \in \Theta} \min_{j=1,2} \Sigma_{gjj,n}^{11}(\theta) > 0$ .

(ii) Let  $M_{g,n}^{11}(\theta) = \text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-1/2}$ . Then

$$\begin{aligned}
\|R_{g,n}^{11}(\theta)\| &= \|M_{g,n}^{11}(\theta)\Sigma_{g,n}^{11}(\theta)M_{g,n}^{11}(\theta)\| \\
&\leq \|M_{g,n}^{11}(\theta)\|^2\|\Sigma_{g,n}^{11}(\theta)\| \\
&= \|M_{g,n}^{11}(\theta)\|^2\|\tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{\Omega}_{g,n}^{so'}(\theta)\| \\
&\leq \|M_{g,n}^{11}(\theta)\|^2(\|\Omega_{g,n}^{ss}(\theta)\| + \|\Omega_{g,n}^{so}(\theta)\|^2\|\Omega_{g,n}^{oo-1}(\theta)\|) \\
&< \infty
\end{aligned} \tag{I.1}$$

uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$  because of Lemma E.1 and the first part of this lemma, which guarantees that  $\|M_{g,n}^{11}(\theta)\|$  is uniformly bounded. By Lemma D.7,

$$\begin{aligned}
\left\| \frac{\partial |R_{g,n}^{11}(\theta)|}{\partial \theta} \right\| &= \left\| |R_{g,n}^{11}(\theta)| \left( \frac{\partial \text{vec } R_{g,n}^{11}(\theta)}{\partial \theta'} \right)' \text{vec } R_{g,n}^{11^{-1}}(\theta) \right\| \\
&\leq |R_{g,n}^{11}(\theta)| \left\| \frac{\partial \text{vec } R_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \|R_{g,n}^{11^{-1}}(\theta)\|.
\end{aligned} \tag{I.2}$$

The uniform boundedness of  $\|R_{g,n}^{11}(\theta)\|$  implies that  $|R_{g,n}^{11}(\theta)|$  is uniformly bounded as well. By the first part of the proof,  $\inf_{n,g} \inf_{\theta \in \Theta} |R_{g,n}^{11}(\theta)| = \inf_{n,g} \inf_{\theta \in \Theta} (1 - \rho_{g,n}^{11^2}(\theta)) > 0$ . Thus, the last term in (I.2) is uniformly bounded by noticing that  $\|R_{g,n}^{11^{-1}}(\theta)\| = |R_{g,n}^{11}(\theta)|^{-1}\|R_{g,n}^{11}(\theta)\|$ .

It remains to show that the second term in (I.2) is uniformly bounded:

$$\begin{aligned}
\left\| \frac{\partial \text{vec } R_{g,n}^{11}(\theta)}{\partial \theta'} \right\| &= \left\| \frac{\partial \text{vec}(M_{g,n}^{11}(\theta)\Sigma_{g,n}^{11}(\theta)M_{g,n}^{11}(\theta))}{\partial \theta'} \right\| \\
&= \left\| \left[ (M_{g,n}^{11}(\theta)\Sigma_{g,n}^{11}(\theta) \otimes I_2) + (I_2 \otimes M_{g,n}^{11}(\theta)\Sigma_{g,n}^{11}(\theta)) \right] \frac{\partial \text{vec } M_{g,n}^{11}(\theta)}{\partial \theta'} + (M_{g,n}^{11}(\theta) \otimes M_{g,n}^{11}(\theta)) \frac{\partial \text{vec } \Sigma_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \\
&\leq (\|M_{g,n}^{11}(\theta)\Sigma_{g,n}^{11}(\theta) \otimes I_2\| + \|I_2 \otimes M_{g,n}^{11}(\theta)\Sigma_{g,n}^{11}(\theta)\|) \left\| \frac{\partial \text{vec } M_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \\
&\quad + \|M_{g,n}^{11}(\theta) \otimes M_{g,n}^{11}(\theta)\| \left\| \frac{\partial \text{vec } \Sigma_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \\
&\leq 2\sqrt{2}\|M_{g,n}^{11}(\theta)\|\|\Sigma_{g,n}^{11}(\theta)\| \left\| \frac{\partial \text{vec } M_{g,n}^{11}(\theta)}{\partial \theta'} \right\| + \|M_{g,n}^{11}(\theta)\|^2 \left\| \frac{\partial \text{vec } \Sigma_{g,n}^{11}(\theta)}{\partial \theta'} \right\|.
\end{aligned}$$

The norm of  $M_{g,n}^{11}(\theta)$  is uniformly bounded due to the first part of this lemma, whereas the boundedness of  $\|\Sigma_{g,n}^{11}(\theta)\|$

is implied by the proof of the boundedness of  $\|R_{g,n}^{11}(\theta)\|$  in (I.1). Note that

$$\left\| \frac{\partial \text{vec } M_{g,n}^{11}(\theta)}{\partial \theta'} \right\| = \sqrt{\left\| \frac{\partial M_{g,n}^{11}(\theta)}{\partial \lambda^s} \right\|^2 + \left\| \frac{\partial M_{g,n}^{11}(\theta)}{\partial \lambda^o} \right\|^2 + \left\| \frac{\partial M_{g,n}^{11}(\theta)}{\partial \rho} \right\|^2 + \left\| \frac{\partial M_{g,n}^{11}(\theta)}{\partial \sigma^2} \right\|^2}.$$

We will show that the first term is uniformly bounded, while the uniform boundedness of the other terms can be established in a similar way:

$$\left\| \frac{\partial M_{g,n}^{11}(\theta)}{\partial \lambda^s} \right\| = \left\| -\frac{1}{2} \text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-3/2} \text{Diag} \left( \frac{\partial \Sigma_{g,n}^{11}(\theta)}{\partial \lambda^s} \right) \right\| \leq \|\text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-3/2}\| \left\| \frac{\partial \Sigma_{g,n}^{11}(\theta)}{\partial \lambda^s} \right\|.$$

The first term on the right hand side is uniformly bounded due to the first part of the lemma. Regarding the second term, note that  $\|\partial \Sigma_{g,n}^{11}(\theta)/\partial \lambda^s\| = \|\partial(\tilde{\Omega}_{g,n}^{ss}(\theta) - \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)\tilde{\Omega}_{g,n}^{so'}(\theta))/\partial \lambda^s\|$ ; it is uniformly bounded by the triangle inequality, the product rule, sub-multiplicity of matrix norms, and Lemma E.1. The boundedness of  $\|\partial \text{vec } \Sigma_{g,n}^{11}(\theta)/\partial \theta'\|$  can be established similarly. It concludes the proof that the second term in (I.2) is uniformly bounded.

Finally,

$$\left\| \frac{\partial \text{vec } R_{g,n}^{11^{-1}}(\theta)}{\partial \theta'} \right\| = \left\| \left( R_{g,n}^{11^{-1}}(\theta) \otimes R_{g,n}^{11^{-1}}(\theta) \right) \frac{\partial \text{vec } R_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \leq \|R_{g,n}^{11^{-1}}(\theta)\|^2 \left\| \frac{\partial \text{vec } R_{g,n}^{11}(\theta)}{\partial \theta'} \right\|$$

is uniformly bounded in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$  by the previous results of this proof.  $\square$

**Proof of Lemma E.3.** Employing the equivalence of vector norms on finite dimensional vector spaces and Loève's  $c_r$ -inequality, it follows for some constant  $C_1 > 0$  that uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} \|S_{g,n}^b(\lambda^b) X_n^b \beta^b\| \right]^p &\leq C_1 E \left[ \sup_{\theta \in \Theta} |S_{g^1,n}^b(\lambda^b) X_n^b \beta^b| + \sup_{\theta \in \Theta} |S_{g^2,n}^b(\lambda^b) X_n^b \beta^b| \right]^p \\ &\leq 2^{p-1} C_1 \left( E \left[ \sup_{\theta \in \Theta} |S_{g^1,n}^b(\lambda^b) X_n^b \beta^b| \right]^p + E \left[ \sup_{\theta \in \Theta} |S_{g^2,n}^b(\lambda^b) X_n^b \beta^b| \right]^p \right) \\ &\leq 2^p C_1 \sup_{n,i} E \left[ \sup_{\theta \in \Theta} |S_{i,n}^b(\lambda^b) X_n^b \beta^b| \right]^p. \end{aligned} \quad (\text{I.3})$$

We proceed with the last term in (I.3): uniformly in  $n \in \mathbb{N}$  and  $i \in \{1, 2, \dots, 2n\}$ ,

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} |S_{i,n}^b(\lambda^b) X_n^b \beta^b| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left| \sum_{j=1}^{2n} S_{ij,n}^b(\lambda^b) X_{j,n}^b \beta^b \right| \right]^p \\
&\leq E \left[ \sup_{\theta \in \Theta} \sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)| \|X_{j,n}^b\| \|\beta^b\| \right]^p \\
&\leq \sup_{\theta \in \Theta} \|\beta^b\|^p \sup_{\theta \in \Theta} \left( \sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)| \right)^p E \left[ \sup_{\theta \in \Theta} \frac{\sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)| \|X_{j,n}^b\|}{\sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)|} \right]^p \quad (\text{I.4}) \\
&\leq \sup_{\theta \in \Theta} \|\beta^b\|^p \sup_{\theta \in \Theta} \left( \sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)| \right)^p \sup_{\theta \in \Theta} \frac{\sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)| E \|X_{j,n}^b\|^p}{\sum_{j=1}^{2n} |S_{ij,n}^b(\lambda^b)|} \\
&\leq \sup_{\theta \in \Theta} \|\beta^b\|^p \sup_n \sup_{\theta \in \Theta} \|S_n^b(\lambda^b)\|_\infty^p \sup_{n,i} E \|X_{i,n}^b\|^p \\
&< \infty,
\end{aligned}$$

where the third inequality follows by Jensen's inequality for convex functions. The conclusion is implied by Assumptions 1(ii), 4(ii), and 7.

Next,

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \|y_{g,n}^o - S_{g,n}^o(\lambda^o) X_n^o \beta^o\| \right]^p \\
&\leq 2^{p-1} \left( E \|y_{g,n}^o\|^p + E \left[ \sup_{\theta \in \Theta} \|S_{g,n}^o(\lambda^o) X_n^o \beta^o\| \right]^p \right), \quad (\text{I.5})
\end{aligned}$$

where the inequality follows by the triangle and Loève's  $c_r$ -inequalities. Given that we have already shown that the second term in (I.5) is uniformly bounded, it is enough to establish that  $\sup_{n,i} E |y_{i,n}^o|^p < \infty$ :

$$\begin{aligned}
E |y_{i,n}^o|^p &\leq E [|y_{i,n}^{*o}|^p | y_{i,n}^s = 1] = E [|S_{i,n}^o(\lambda_0^o) X_n^o \beta_0^o + \varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1] \\
&\leq 2^{p-1} (E [|S_{i,n}^o(\lambda_0^o) X_n^o \beta_0^o|^p | y_{i,n}^s = 1] + E [|\varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1]).
\end{aligned}$$

Now we show that each term is bounded. In a similar way as in (I.4),

$$E [|S_{i,n}^o(\lambda_0^o) X_n^o \beta_0^o|^p | y_{i,n}^s = 1] \leq \|\beta_0^o\|^p \sup_n \|S_n^o(\lambda_0^o)\|_\infty^p \sup_{n,i} E [|X_{i,n}^o|^p | y_{i,n}^s = 1] < \infty$$

by Assumptions 1(ii), 4(ii), and 7. By the law of iterated expectations,

$$E [|\varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1] = E [E [|\varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1, X_n^s] | y_{i,n}^s = 1]. \quad (\text{I.6})$$



Firstly, we will find the inner expectation by applying the law of iterated expectations once more:

$$E[|\varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1, X_n^s] = E[E[|\varepsilon_{i,n}^o(\lambda_0^o)|^p | \varepsilon_{i,n}^s(\lambda_0^s), X_n^s] | y_{i,n}^s = 1, X_n^s] = E[E[|\varepsilon_{i,n}^o(\lambda_0^o)|^p | \varepsilon_{i,n}^s(\lambda_0^s)] | y_{i,n}^s = 1, X_n^s], \quad (\text{I.7})$$

where the last equality follows by Assumption 2(ii).

Note that  $(\varepsilon_{i,n}^o(\lambda_0^o), \varepsilon_{i,n}^s(\lambda_0^s))' \sim \mathcal{N}(0, [\Omega_{ii,n}^{oo}(\theta_0) \Omega_{ii,n}^{so}(\theta_0); \Omega_{ii,n}^{so}(\theta_0) \Omega_{ii,n}^{ss}(\theta_0)])$ . Thus,  $\varepsilon_{i,n}^o(\lambda_0^o) | \varepsilon_{i,n}^s(\lambda_0^s) \sim \mathcal{N}(\tilde{\mu}_{i,n}(\theta_0), \tilde{\sigma}_{i,n}^2(\theta_0))$  with  $\tilde{\mu}_{i,n}(\theta_0) = \sqrt{\Omega_{ii,n}^{oo}(\theta_0)/\Omega_{ii,n}^{ss}(\theta_0)} \rho_{i,n}^{so}(\theta_0) \varepsilon_{i,n}^s(\lambda_0^s)$  and  $\tilde{\sigma}_{i,n}^2(\theta_0) = (1 - \rho_{i,n}^{so}(\theta_0)^2) \Omega_{ii,n}^{oo}(\theta_0)$ , where  $\rho_{i,n}^{so}(\theta_0)$  is the correlation coefficient of  $\varepsilon_{i,n}^s(\lambda_0^s)$  and  $\varepsilon_{i,n}^o(\lambda_0^o)$ . Hence, the inner expectation in (I.7) can be bounded by

$$\begin{aligned} E[|\varepsilon_{i,n}^o(\lambda_0^o)|^p | \varepsilon_{i,n}^s(\lambda_0^s)] &= E[|\varepsilon_{i,n}^o(\lambda_0^o) - \tilde{\mu}_{i,n}(\theta_0) + \tilde{\mu}_{i,n}(\theta_0)|^p | \varepsilon_{i,n}^s(\lambda_0^s)] \\ &\leq 2^{p-1} (E[|\varepsilon_{i,n}^o(\lambda_0^o) - \tilde{\mu}_{i,n}(\theta_0)|^p | \varepsilon_{i,n}^s(\lambda_0^s)] + E[|\tilde{\mu}_{i,n}(\theta_0)|^p | \varepsilon_{i,n}^s(\lambda_0^s)]) \\ &= 2^{p-1} \left( \tilde{\sigma}_{i,n}^p(\theta_0) 2^{p/2} \Gamma\left(\frac{p+1}{2}\right) / \sqrt{\pi} + |\tilde{\mu}_{i,n}(\theta_0)|^p \right) \\ &\leq C_2 + C_3 |\varepsilon_{i,n}^s(\lambda_0^s)|^p \end{aligned}$$

for some constants  $C_2, C_3 > 0$ , where  $\Gamma(\cdot)$  is the Gamma function. The second equality is implied by the following fact: if  $X \sim \mathcal{N}(0, \sigma^2)$ , then for any  $p \in (-1, +\infty)$ ,  $E[|X|^p] = \sigma^p 2^{p/2} \Gamma((p+1)/2) / \sqrt{\pi}$  (Kamat, 1953). The conclusion follows by noticing that Lemma E.1 implies the uniform boundedness from zero of  $\Omega_{ii,n}^{ss}(\theta_0)$  and the uniform boundedness of  $\Omega_{ii,n}^{oo}(\theta_0)$  which is implied by the uniform boundedness of  $\|\Omega_{g,n}^{oo}(\theta_0)\|$ . Thus, the expectation in (I.7) becomes

$$\begin{aligned} E[|\varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1, X_n^s] &\leq C_2 + C_3 E[|\varepsilon_{i,n}^s(\lambda_0^s)|^p - \varepsilon_{i,n}^s(\lambda_0^s) < S_{i,n}^s(\lambda_0^s) X_n^s \beta_0^s, X_n^s] \\ &= C_2 + C_3 (\Omega_{ii,n}^{ss}(\theta_0))^{p/2} E \left[ \left| \frac{\varepsilon_{i,n}^s(\lambda_0^s)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}} \right|^p - \frac{\varepsilon_{i,n}^s(\lambda_0^s)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}} < \frac{S_{i,n}^s(\lambda_0^s) X_n^s \beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}, X_n^s \right] \\ &\leq C_2 + C_3 (\Omega_{ii,n}^{ss}(\theta_0))^{p/2} \left( E \left[ \left| \frac{\varepsilon_{i,n}^s(\lambda_0^s)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}} \right|^r - \frac{\varepsilon_{i,n}^s(\lambda_0^s)}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}} < \frac{S_{i,n}^s(\lambda_0^s) X_n^s \beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}}, X_n^s \right] \right)^{p/r} \\ &= C_2 + C_3 (\Omega_{ii,n}^{ss}(\theta_0))^{p/2} \vartheta_r^{p/r}(m_{i,n}(\theta_0)) \\ &\leq C_2 + C_4 \vartheta_r^{p/r}(m_{i,n}(\theta_0)) \end{aligned}$$

for some constant  $C_4 > 0$ , where  $r$  is the smallest integer at least as large as  $p$ ,  $m_{i,n}(\theta_0) = S_{i,n}^s(\lambda_0^s) X_n^s \beta_0^s / \sqrt{\Omega_{ii,n}^{ss}(\theta_0)}$ , and  $\vartheta_r(m_{i,n}(\theta_0)) \leq |m_{i,n}(\theta_0)|^{r-1} \phi(m_{i,n}(\theta_0)) / \Phi(m_{i,n}(\theta_0)) + (r-1) \vartheta_{r-2}(m_{i,n}(\theta_0))$  for  $r \geq 2$  with  $\vartheta_1(m_{i,n}(\theta_0)) \leq \phi(m_{i,n}(\theta_0)) / \Phi(m_{i,n}(\theta_0)) + C_5$  and  $\vartheta_0 = 1$ , for some constant  $C_5 > 0$ . The second inequality is implied by Hölder's

inequality, the second equality follows by Lemma G.1, whereas the last inequality follows from the uniform boundedness of  $\Omega_{ii,n}^{ss}(\theta_0)$ , which is implied by the uniform boundedness of  $\|\Omega_{g,n}^{ss}(\theta)\|$  established in Lemma E.1. Consider  $r \geq 2$ . It follows from (I.6) that

$$\begin{aligned} E[|\varepsilon_{i,n}^o(\lambda_0^o)|^p | y_{i,n}^s = 1] &\leq C_2 + C_4 E[\vartheta_r^{p/r}(m_{i,n}(\theta_0)) | y_{i,n}^s = 1] \\ &\leq C_2 + C_4 E \left[ |m_{i,n}(\theta_0)|^{r-1} \frac{\phi(m_{i,n}(\theta_0))}{\Phi(m_{i,n}(\theta_0))} + (r-1)\vartheta_{r-2}(m_{i,n}(\theta_0)) \middle| y_{i,n}^s = 1 \right]^{p/r} \\ &\leq C_2 + C_4 \left( E \left[ |m_{i,n}(\theta_0)|^{r-1} \frac{\phi(m_{i,n}(\theta_0))}{\Phi(m_{i,n}(\theta_0))} \middle| y_{i,n}^s = 1 \right]^{p/r} + (r-1)^{p/r} E[\vartheta_{r-2}(m_{i,n}(\theta_0)) | y_{i,n}^s = 1]^{p/r} \right). \end{aligned} \quad (\text{I.8})$$

From the proof of Lemma A.9 by Xu and Lee (2015), it follows that  $\phi(x)/\Phi(x) \leq 2(|x| + C_6)$ , for some constant  $C_6 > 0$ . Hence,

$$\begin{aligned} E \left[ |m_{i,n}(\theta_0)|^{r-1} \frac{\phi(m_{i,n}(\theta_0))}{\Phi(m_{i,n}(\theta_0))} \middle| y_{i,n}^s = 1 \right]^{p/r} &\leq 2^{p/r} E \left[ |m_{i,n}(\theta_0)|^{r-1} (|m_{i,n}(\theta_0)| + C_6) \middle| y_{i,n}^s = 1 \right]^{p/r} \\ &\leq 2^{p/r} \left( E[|m_{i,n}(\theta_0)|^p | y_{i,n}^s = 1] + C_6^{p/r} E[|m_{i,n}(\theta_0)|^{p(r-1)/r} | y_{i,n}^s = 1] \right), \end{aligned}$$

where the last inequality follows by Loève's  $c_r$ -inequality. In order to show that the first term in (I.8) is uniformly bounded, by Hölder's inequality it is enough to establish that  $E[|m_{i,n}(\theta_0)|^p | y_{i,n}^s = 1]$  is uniformly bounded. In the same way as in (I.4),

$$\begin{aligned} E[|m_{i,n}(\theta_0)|^p | y_{i,n}^s = 1] &= E \left[ \left| \frac{S_{i,n}^s(\lambda_0^s) X_n^s \beta_0^s}{\sqrt{\Omega_{ii,n}^{ss}(\theta_0)}} \right|^p \middle| y_{i,n}^s = 1 \right] \\ &\leq \sup_{n,i} \Omega_{ii,n}^{ss-p/2}(\theta_0) \|\beta_0^s\|^p \sup_n \|S_n^s(\lambda_0^s)\|_\infty^p \sup_{n,i} E[|X_{i,n}^s|^p | y_{i,n}^s = 1] < \infty, \end{aligned}$$

where the conclusion is implied by Assumptions 1(ii), 4(ii), and 7 and Lemma E.1.

It is easy to show using recursion and Hölder's inequality that the second term in (I.8) is uniformly bounded if  $E[|m_{i,n}(\theta_0)|^p | y_{i,n}^s = 1]$  is uniformly bounded. This condition is sufficient for the case when  $r = 1$  as well. It completes the proof that  $\sup_{\theta \in \Theta} \|z_{g,n}(\theta)\|$  is uniformly  $L_p$ -bounded.

We continue by showing that  $\sup_{\theta \in \Theta} \|\mu_{g,n}^{11}(\theta)\|$  is uniformly  $L_p$ -bounded:

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} \|\mu_{g,n}^{11}(\theta)\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \|\tilde{\Omega}_{g,n}^{so}(\theta) \Omega_{g,n}^{oo-1}(\theta) z_{g,n}(\theta)\| \right]^p \\ &\leq \sup_{\theta \in \Theta} (\|\Omega_{g,n}^{so}(\theta)\| \|\Omega_{g,n}^{oo-1}(\theta)\|)^p E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^p < \infty \end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , where the conclusion is implied by Lemma E.1 and the previous results of this proof.

In the same way as in (I.3),  $\sup_{\theta \in \Theta} \|v_{g,n}^{11}(\theta)\|$  is uniformly  $L_p$ -bounded if  $\sup_{\theta \in \Theta} |v_{gj,n}^{11}(\theta)|$  is uniformly bounded for  $j = 1, 2$ :

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} |v_{gj,n}^{11}(\theta)| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left| \frac{\tilde{S}_{gj,n}^s(\lambda^s) X_n^s \beta^s - \mu_{gj,n}^{11}(\theta)}{\sqrt{\Sigma_{gjj,n}^{11}(\theta)}} \right| \right]^p \\ &\leq 2^{p-1} \sup_{\theta \in \Theta} \Sigma_{gjj,n}^{11-p/2}(\theta) \left( E \left[ \sup_{\theta \in \Theta} |S_{gj,n}^s(\lambda^s) X_n^s \beta^s| \right]^p + E \left[ \sup_{\theta \in \Theta} |\mu_{gj,n}^{11}(\theta)| \right]^p \right) < \infty \end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ , where the conclusion follows by the previous results of this proof and Lemma E.2.  $\square$

**Proof of Lemma E.4.** Applying Loève's  $c_r$ -inequality twice leads to the following bound:

$$\begin{aligned} E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial z_{g,n}(\theta)}{\partial \theta'} \right\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial (y_{g,n}^o - S_{g,n}^o(\lambda^o) X_n^o \beta^o)}{\partial \theta'} \right\| \right]^p \\ &= E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial (S_{g,n}^o(\lambda^o) X_n^o \beta^o)}{\partial \theta'} \right\| \right]^p \\ &= E \left[ \sup_{\theta \in \Theta} \sqrt{\left\| \frac{\partial (S_{g,n}^o(\lambda^o) X_n^o \beta^o)}{\partial \lambda^o} \right\|^2 + \left\| \frac{\partial (S_{g,n}^o(\lambda^o) X_n^o \beta^o)}{\partial \beta^{o'}} \right\|^2} \right]^p \quad (\text{I.9}) \\ &\leq 2^{p-1} \left( E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial (S_{g,n}^o(\lambda^o) X_n^o \beta^o)}{\partial \lambda^o} \right\| \right]^p + E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial (S_{g,n}^o(\lambda^o) X_n^o \beta^o)}{\partial \beta^{o'}} \right\| \right]^p \right) \\ &= 2^{p-1} \left( E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial S_{g,n}^o(\lambda^o)}{\partial \lambda^o} X_n^o \beta^o \right\| \right]^p + E \left[ \sup_{\theta \in \Theta} \|S_{g,n}^o(\lambda^o) X_n^o\| \right]^p \right). \end{aligned}$$

The uniform boundedness of the second term can be proven in the same way as the uniform boundedness of  $E[\sup_{\theta \in \Theta} \|S_{g,n}^b(\lambda^b) X_n^b \beta^b\|]^p$  in Lemma E.3. In the same way as in (I.3) and (I.4), we can show that the first term is uniformly bounded if

$$\sup_{\theta \in \Theta} \|\beta^o\|^p \sup_n \sup_{\theta \in \Theta} \left\| \frac{\partial S_n^o(\lambda^o)}{\partial \lambda^o} \right\|_\infty^p \sup_{n,i} E \|X_{i,n}^o\|^p < \infty. \quad (\text{I.10})$$

The first and the last terms in (I.10) are bounded by Assumptions 7 and 4(ii), respectively. The second term in (I.10) is uniformly bounded because

$$\left\| \frac{\partial S_n^o(\lambda^o)}{\partial \lambda^o} \right\|_\infty = \left\| \frac{\partial (I_{2n} - \lambda^o W_n^o)^{-1}}{\partial \lambda^o} \right\|_\infty = \|(I_{2n} - \lambda^o W_n^o)^{-1} W_n^o (I_{2n} - \lambda^o W_n^o)^{-1}\|_\infty < \infty,$$

where the result follows from the sub-multiplicativity of matrix norms and Assumption 1(ii).

Next, let  $M_{g,n}^{11}(\theta) = \text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-1/2}$ . Then

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial v_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial (M_{g,n}^{11}(\theta)(q_{g,n}(\theta) - \mu_{g,n}^{11}(\theta)))}{\partial \theta'} \right\| \right]^p \\
&= E \left[ \sup_{\theta \in \Theta} \left\| ((q_{g,n}(\theta) - \mu_{g,n}^{11}(\theta))' \otimes I_2) \frac{\partial \text{vec } M_{g,n}^{11}(\theta)}{\partial \theta'} + M_{g,n}^{11}(\theta) \left( \frac{\partial q_{g,n}(\theta)}{\partial \theta'} - \frac{\partial \mu_{g,n}^{11}(\theta)}{\partial \theta'} \right) \right\| \right]^p \\
&\leq 2^{p-1} \left( \sup_{\theta \in \Theta} \left\| \frac{\partial \text{vec } M_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right)^p E \left[ \sup_{\theta \in \Theta} \|(q_{g,n}(\theta) - \mu_{g,n}^{11}(\theta))' \otimes I_2\| \right]^p \\
&\quad + \sup_{\theta \in \Theta} \|M_{g,n}^{11}(\theta)\|^p E \left[ \sup_{\theta \in \Theta} \left( \left\| \frac{\partial q_{g,n}(\theta)}{\partial \theta'} \right\| + \left\| \frac{\partial \mu_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right) \right]^p
\end{aligned} \tag{I.11}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$ . We have already shown in the proof of Lemma E.2 that the norms of  $\partial \text{vec } M_{g,n}^{11}(\theta)/\partial \theta'$  and  $M_{g,n}^{11}(\theta)$  are uniformly bounded. By Lemma D.8,

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} \|(q_{g,n}(\theta) - \mu_{g,n}^{11}(\theta))' \otimes I_2\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \|q_{g,n}(\theta) - \mu_{g,n}^{11}(\theta)\| \|I_2\| \right]^p \\
&\leq 2^{3p/2-1} \left( E \left[ \sup_{\theta \in \Theta} \|q_{g,n}(\theta)\| \right]^p + E \left[ \sup_{\theta \in \Theta} \|\mu_{g,n}^{11}(\theta)\| \right]^p \right) \\
&< \infty
\end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  by Lemma E.3 because  $\|q_{g,n}(\theta)\| = \|\tilde{S}_{g,n}^s(\lambda^s)X_n^s\beta^s\| = \|S_{g,n}^s(\lambda^s)X_n^s\beta^s\|$ . It remains to show that the last term in (I.11) is uniformly bounded. By Loève's  $c_r$ -inequality,

$$E \left[ \sup_{\theta \in \Theta} \left( \left\| \frac{\partial q_{g,n}(\theta)}{\partial \theta'} \right\| + \left\| \frac{\partial \mu_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right) \right]^p \leq 2^{p-1} \left( E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial q_{g,n}(\theta)}{\partial \theta'} \right\| \right]^p + E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial \mu_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right]^p \right).$$

We can show that the first term is uniformly bounded in the same way as we proved earlier in this proof that  $\sup_{n,g} E[\sup_{\theta \in \Theta} \|\partial(S_{g,n}^o(\lambda^o)X_n^o\beta^o)/\partial \theta'\|]^p < \infty$ . For the second term, note that

$$\begin{aligned}
E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial \mu_{g,n}^{11}(\theta)}{\partial \theta'} \right\| \right]^p &= E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial (\tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta))}{\partial \theta'} \right\| \right]^p \\
&= E \left[ \sup_{\theta \in \Theta} \left\| (z'_{g,n}(\theta) \otimes I_2) \frac{\partial \text{vec}(\tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta))}{\partial \theta'} + \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta) \frac{\partial z_{g,n}(\theta)}{\partial \theta'} \right\| \right]^p \\
&\leq 2^{p-1} \left( \sup_{\theta \in \Theta} \left\| \frac{\partial \text{vec}(\tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta))}{\partial \theta'} \right\| \right)^p 2^{p/2} E \left[ \sup_{\theta \in \Theta} \|z_{g,n}(\theta)\| \right]^p \\
&\quad + \sup_{\theta \in \Theta} (\|\Omega_{g,n}^{so}(\theta)\| \|\Omega_{g,n}^{oo-1}(\theta)\|)^p E \left[ \sup_{\theta \in \Theta} \left\| \frac{\partial z_{g,n}(\theta)}{\partial \theta'} \right\| \right]^p
\end{aligned}$$

is uniformly bounded. The conclusion then follows by Lemmas E.1 and E.3, the previous results of this lemma, and by noticing that

$$\begin{aligned} \left\| \frac{\partial \text{vec}(\tilde{\Omega}_{g,n}^{so}(\theta) \Omega_{g,n}^{oo-1}(\theta))}{\partial \theta'} \right\| &= \left\| (\Omega_{g,n}^{oo-1}(\theta) \otimes I_2) \frac{\partial \text{vec} \tilde{\Omega}_{g,n}^{so}(\theta)}{\partial \theta'} + (I_2 \otimes \tilde{\Omega}_{g,n}^{so}(\theta)) \frac{\partial \text{vec} \Omega_{g,n}^{oo-1}(\theta)}{\partial \theta'} \right\| \\ &\leq \sqrt{2} \left( \left\| \Omega_{g,n}^{oo-1}(\theta) \right\| \left\| \frac{\partial \text{vec} \Omega_{g,n}^{so}(\theta)}{\partial \theta'} \right\| + \left\| \Omega_{g,n}^{so}(\theta) \right\| \left\| \frac{\partial \text{vec} \Omega_{g,n}^{oo-1}(\theta)}{\partial \theta'} \right\| \right) \\ &< \infty \end{aligned}$$

uniformly in  $n \in \mathbb{N}$  and  $g \in \mathcal{G}_n$  by Lemma E.1.  $\square$

**Proof of Lemma E.5.** We start with establishing the uniform  $L_p$ -NED,  $p \in \{2, 4\}$ , property for  $\{y_{g,n}^{*b}\}_{g \in \mathcal{G}_n}$ ,  $b \in \{s, o\}$ , which will be needed later in the proof. Using the definition of NED and the conditional Jensen's inequality, it follows

$$\begin{aligned} \|y_{g,n}^{*b} - E[y_{g,n}^{*b} | \mathcal{F}_{g,n}(s)]\|_p &= \|S_{g,n}^b(\lambda_0^b) (X_n^b \beta_0^b + u_n^b - E[X_n^b \beta_0^b + u_n^b | \mathcal{F}_{g,n}(s)])\|_p \\ &= \left\| \sum_{\dot{g} \in \mathcal{G}_n} S_{g\dot{g},n}^b(\lambda_0^b) (X_{\dot{g},n}^b \beta_0^b + u_{\dot{g},n}^b - E[X_{\dot{g},n}^b \beta_0^b + u_{\dot{g},n}^b | \mathcal{F}_{g,n}(s)]) \right\|_p \\ &= \left\| \sum_{\dot{g}: d(g, \dot{g}) > s} S_{g\dot{g},n}^b(\lambda_0^b) (X_{\dot{g},n}^b \beta_0^b + u_{\dot{g},n}^b - E[X_{\dot{g},n}^b \beta_0^b + u_{\dot{g},n}^b | \mathcal{F}_{g,n}(s)]) \right\|_p \\ &\leq \sum_{\dot{g}: d(g, \dot{g}) > s} \|S_{g\dot{g},n}^b(\lambda_0^b)\| (\|X_{\dot{g},n}^b \beta_0^b + u_{\dot{g},n}^b\|_p + \|E[X_{\dot{g},n}^b \beta_0^b + u_{\dot{g},n}^b | \mathcal{F}_{g,n}(s)]\|_p) \\ &\leq 2 \sum_{\dot{g}: d(g, \dot{g}) > s} \|S_{g\dot{g},n}^b(\lambda_0^b)\| (\|X_{\dot{g},n}^b\|_p \|\beta_0^b\| + \|u_{\dot{g},n}^b\|_p) \\ &\leq 2 \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda^b)\| \cdot \left( \sup_{n,g} \|X_{\dot{g},n}^b\|_p \|\beta_0^b\| + \sup_{n,g} \|u_{\dot{g},n}^b\|_p \right) \\ &\quad \times \frac{\sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g}: d(g, \dot{g}) > s} \|S_{g\dot{g},n}^b(\lambda^b)\|}{\sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda^b)\|} \\ &\leq t^{y^{*b}} \psi(s), \end{aligned}$$

where  $t^{y^{*b}} = 2 \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda^b)\| \left( \sup_{n,g} \|X_{\dot{g},n}^b\|_p \|\beta_0^b\| + \sup_{n,g} \|u_{\dot{g},n}^b\|_p \right)$  and  $\psi(s) = \max\{\psi^s(s), \psi^o(s)\} \leq 1$  with  $\psi^b(s) = \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g}: d(g, \dot{g}) > s} \|S_{g\dot{g},n}^b(\lambda^b)\| / \sup_{n,g} \sup_{\theta \in \Theta} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda^b)\|$ ,  $b \in \{s, o\}$ . The first and second inequalities follow by Minkowski's and the conditional Jensen's inequalities, respectively. Given Assumption 7,  $t^{y^{*b}}$  is bounded provided that  $E\|X_{g,n}^b\|^p$ ,  $E\|u_{g,n}^b\|^p$ , and  $\sum_{\dot{g} \in \mathcal{G}_n} \|S_{g\dot{g},n}^b(\lambda^b)\|$  are uniformly bounded. Since  $p \in \{2, 4\}$ , by Liapunov's inequality it is enough to establish the results for  $p = 4$ . Notice that  $E\|X_{g,n}^b\|^4 =$

$E [\|X_{g1,n}^b\|^2 + \|X_{g2,n}^b\|^2]^2 \leq \sup_{n,i} 4E\|X_{i,n}^b\|^4 < \infty$  by Assumption 4(ii). Because a normal distribution has infinitely many moments and  $\sup_{n,g} E\|u_{g,n}^b\|^4 \leq \sup_{n,i} 4E|u_{i,n}^b|^4$ , Assumption 2(i) implies that  $\sup_{n,g} E\|u_{g,n}^b\|^4 < \infty$ , whereas equivalence of matrix norms on finite dimensional matrix spaces implies that uniformly in  $n \in \mathbb{N}$ ,  $g \in \mathcal{G}_n$ , and  $\theta \in \Theta$

$$\begin{aligned} \sum_{\dot{g} \in \mathcal{G}_n} \|S_{\dot{g}g,n}^b(\lambda^b)\| &\leq C_1 \sum_{\dot{g} \in \mathcal{G}_n} \|S_{\dot{g}g,n}^b(\lambda^b)\|_\infty \leq C_1 \sum_{\dot{g} \in \mathcal{G}_n} (\|S_{g1\dot{g},n}^b(\lambda^b)\|_\infty + \|S_{g2\dot{g},n}^b(\lambda^b)\|_\infty) \\ &\leq 2C_1 \|S_{g,n}^b(\lambda^b)\|_\infty \leq 2C_1 \|S_n^b(\lambda^b)\|_\infty < \infty \end{aligned} \quad (\text{I.12})$$

for some constant  $C_1 > 0$  by Assumption 1(ii). Note that by Assumption 5,  $\lim_{s \rightarrow \infty} \psi(s) = 0$ . Thus,  $\{y_{g,n}^{*b}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_4$ - and  $L_2$ -NED random field with NED coefficients  $\psi(s)$ .

Recall that  $d_{g,n}^{11} = \mathbb{1}(y_{g1,n}^s = 1, y_{g2,n}^s = 1) = \mathbb{1}(y_{g1,n}^{*s} > 0) \cdot \mathbb{1}(y_{g2,n}^{*s} > 0)$ . From the proof of Proposition 2 by Xu and Lee (2015), it follows that for some constants  $C_2, C_3 > 0$  and  $j = 1, 2$ ,

$$\|\mathbb{1}(y_{gj,n}^{*s} > 0) - E[\mathbb{1}(y_{gj,n}^{*s} > 0)|\mathcal{F}_{g,n}(s)]\|_2 \leq (1 + C_2) \|y_{gj,n}^{*s} - E[y_{gj,n}^{*s}|\mathcal{F}_{g,n}(s)]\|_2^{1/3} \leq (1 + C_2) C_3 \psi^{1/3}(s),$$

where the last inequality follows by Lemma G.2 and the fact that  $\{y_{g,n}^{*s}\}_{g \in \mathcal{G}_n}$  is uniformly  $L_2$ -NED with NED coefficients  $\psi(s)$ . Since  $|\mathbb{1}(y_{gj,n}^{*s} > 0) - E[\mathbb{1}(y_{gj,n}^{*s} > 0)|\mathcal{F}_{g,n}(s)]|^4 \leq |\mathbb{1}(y_{gj,n}^{*s} > 0) - E[\mathbb{1}(y_{gj,n}^{*s} > 0)|\mathcal{F}_{g,n}(s)]|^2$  implies that  $\|\mathbb{1}(y_{gj,n}^{*s} > 0) - E[\mathbb{1}(y_{gj,n}^{*s} > 0)|\mathcal{F}_{g,n}(s)]\|_4 \leq \|\mathbb{1}(y_{gj,n}^{*s} > 0) - E[\mathbb{1}(y_{gj,n}^{*s} > 0)|\mathcal{F}_{g,n}(s)]\|_2^{1/2}$ ,  $\{\mathbb{1}(y_{gj,n}^{*s} > 0)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_4$ -NED random field with NED coefficients  $\psi^{1/6}(s)$ . Given that  $\{\mathbb{1}(y_{gj,n}^{*s} > 0)\}_{g \in \mathcal{G}_n}$  is uniformly  $L_4$ -bounded, Lemma D.10 implies that  $\{d_{g,n}^{11}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{1/6}(s) + \psi^{1/6}(s) + \psi^{1/3}(s) \leq 3\psi^{1/6}(s)$ .<sup>1</sup>

From the definition,  $z_{g,n}(\theta) = y_{g,n}^o - S_{g,n}^o(\lambda^o)X_n^o\beta^o$ . By Theorem D.4, it is enough to establish the uniform NED property for each term of the summation and find their NED coefficients. Note that  $y_{g,n}^o = (\mathbb{1}(y_{g1,n}^{*s} > 0)y_{g1,n}^{*o}, \mathbb{1}(y_{g2,n}^{*s} > 0)y_{g2,n}^{*o})'$ . Since by Lemma G.2 and the previous results of this proof,  $\{y_{gj,n}^{*s}\}_{g \in \mathcal{G}_n}$  and  $\mathbb{1}(y_{gj,n}^{*s} > 0)$ ,  $j = 1, 2$ , are uniformly  $L_4$ -NED with NED coefficients  $\psi(s)$  and  $\psi^{1/6}(s)$ , respectively, Lemma D.10 implies that  $\{\mathbb{1}(y_{gj,n}^{*s} > 0)y_{gj,n}^{*o}\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{1/6}(s) + \psi(s) + \psi^{7/6}(s) \leq 3\psi^{1/6}(s)$ . By Lemma G.2, the same property is transferred to  $\{y_{g,n}^o\}_{g \in \mathcal{G}_n}$ . It is easy to see from the proof of  $\{y_{g,n}^{*o}\}_{g \in \mathcal{G}_n}$  being an  $L_2$ -NED random field that  $\{S_{g,n}^o(\lambda^o)X_n^o\beta^o\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi(s)$ . Hence,  $\{z_{g,n}(\theta)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{1/6}(s) + \psi(s) \leq 2\psi^{1/6}(s)$ .

Finally, since  $v_{g,n}^{11}(\theta) = \text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-1/2}(q_{g,n}(\theta) - \mu_{g,n}^{11}(\theta))$  and  $\|\text{Diag}(\Sigma_{g,n}^{11}(\theta))^{-1/2}\|$  is uniformly bounded by Lemma E.2, it suffices to establish the uniform NED property for  $q_{g,n}(\theta)$  and  $\mu_{g,n}^{11}(\theta)$  and find their NED

<sup>1</sup>Note that in this case we can treat  $\psi^{1/3}(s)$  as the NED coefficient because 3 can be treated as a part of the NED scaling factor.

coefficients. Recall that  $q_{g,n}(\theta) = \tilde{S}_{g,n}^s(\lambda^s)X_n^s\beta^s$ , thus it is a uniform  $L_2$ -NED random field with NED coefficients  $\psi(s)$ . Moreover,  $\mu_{g,n}^{11}(\theta) = \tilde{\Omega}_{g,n}^{so}(\theta)\Omega_{g,n}^{oo-1}(\theta)z_{g,n}(\theta)$ . Given that the norms of  $\tilde{\Omega}_{g,n}^{so}(\theta)$  and  $\Omega_{g,n}^{oo-1}(\theta)$  are uniformly bounded by Lemma E.1,  $\{\mu_{g,n}^{11}(\theta)\}_{g \in \mathcal{G}_n}$  is a uniform  $L_2$ -NED random field with NED coefficients  $\psi^{1/6}(s)$ . The conclusion follows by Theorem D.4.  $\square$

## References

- Abadir, K. and J. Magnus (2005). *Matrix Algebra*. Cambridge University Press.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- Hashorva, E. and J. Hüsler (2003). On multivariate Gaussian tails. *Annals of the Institute of Statistical Mathematics* 55(3), 507–522.
- Kamat, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications. *Biometrika* 40(1), 20–34.
- Magnus, J. and H. Neudecker (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.
- Muthén, B. (1990). Moments of the censored and truncated bivariate normal distribution. *British Journal of Mathematical and Statistical Psychology* 43(1), 131–143.
- Xu, X. and L.-F. Lee (2015). Maximum likelihood estimation of a spatial autoregressive Tobit model. *Journal of Econometrics* 188(1), 264–280.